

STRUCTURAL INSIGHTS INTO TAIL-
ANCHORED PROTEIN TARGETING BY
GET3

Thesis by

Christian Suloway

In Partial Fulfillment of the Requirements for the

Degree of

Doctor of Philosophy

CALIFORNIA INSTITUTE OF TECHNOLOGY

Pasadena, California

2012

(Defended February 21, 2012)

© 2012

Christian Suloway

All Rights Reserved

ACKNOWLEDGEMENTS

I would like to start by acknowledging all the people who gave me the opportunity to work in science. First, Professor Bridget Carragher and Professor Clint Potter for letting me work in their lab at the Beckmann Institute at the University of Illinois at Urbana-Champaign as an undergraduate. They gave me the opportunity to continue work in their lab when they moved to The Scripps Research Institute and then gave their generous support when I applied for graduate school. Without them I would not be where I am today. Thank you to my coworkers at Scripps including Jim Pulokas, Dr. Anchi Cheng, Joel Quispe and Denis Fellmann for making an excellent environment in which to work. My entry into graduate school at Caltech was supported by Professor Grant Jensen who let me work in his lab the year before I started my doctoral work. During this time he went so far as to teach me the basics of biology and structure determination techniques. I am forever in his debt for all the opportunities he has given me.

Next I would like to thank those who helped me in my graduate studies. Thank you to the members of the Clemons Lab past and present: Matt Thornton, Dr. SureshKumar Ramasamy, Justin Chartron, Dr. Axel Müller, Ma'ayan Zaslaver, Dr. Yuko Tsutsui, Dr. Kyoung-Soon Jang, Harry Gristick, Dr. Shiho Tanaka, Amanda Mock, Alice Chen and Dr. Lada Klaić. I would especially like to thank Ma'ayan and Justin for their contributions to the work presented in this dissertation along with undergraduate researchers Alan Deng, Yea-ra Jo and Grace Wu. I received help and advice from other Caltech researchers including Pavle Nickolovski, Dr. Troy Walton and Mike Rome. Faculty members who gave advice and read our manuscripts include Professor Doug Rees, Professor Tom Miller and Professor Shu-ou Shan. My thesis committee of Professor Pamela Bjorkman, Professor Grant Jensen and Professor Rob Phillips has given

me invaluable advice. I would like to give my sincere appreciation to my thesis advisor, Professor William "Bill" Clemons, for the not only the opportunity to work in structural biology but also for his support and advice in science and life.

Thank you to my friend Emilio for numerous discussions and advice on life at Caltech. I would like to thank all of my family members for their support, especially my Mom who gave so much while raising me. Thank you to my wife Tracy and my son Luke for all of your patience and to whom I dedicate this thesis.

ABSTRACT

Translocation of membrane proteins from the point of synthesis to their integration in the membrane is critical to the function of the cell. Tail-anchored (TA) proteins are an important class of membrane proteins with a single transmembrane domain (TMD) close to the carboxyl-terminus. They are defined topologically by having their amino-terminus in the cytosol and their carboxyl-terminus on the exterior side of the membrane. Since the TMD is sequestered by the ribosome during translation, co-translational translocation of TA proteins by the SRP-dependent pathway is not possible. The Guided Entry of Tail-anchored proteins (GET) pathway post-translationally targets TA proteins to the endoplasmic reticulum (ER) membrane. The conserved nucleotide hydrolase Get3 is the central protein in the pathway that specifically binds the TMD of TA proteins to chaperone them from a sorting complex of Get4, Get5, Sgt2 and other chaperones to an ER membrane receptor formed by Get1 and Get2. We have created a model for the mechanism of Get3 TA protein binding coupled to nucleotide state using X-ray crystallography, structural modeling and mutagenesis experiments. We then demonstrate expression, purification and crystallization of complexes of Get3 with TA proteins for structural studies. Finally, we present a crystal structure of a tetrameric archaeal Get3 homologue that forms a central hydrophobic chamber and is capable of binding TA proteins. Using small-angle X-ray scattering, the structure is comparable to a tetrameric fungal Get3 complex with TA protein, which is capable of TA protein membrane integration *in vitro*. This suggests a model in which a tetramer of Get3 binds TA proteins for delivery to the membrane.

TABLE OF CONTENTS

Acknowledgements	iii
Abstract	v
Table of Contents	vi
Introduction	1
Figures.....	19
Chapter 1: A Model for Tail-anchored Protein Binding by Get3	
Introduction	32
Results.....	35
Figures.....	42
Discussion	47
Materials and Methods	50
Acknowledgements	54
Supplemental Figures.....	55
Chapter 2: Toward Crystallization of a Get3 and Tail-anchored Protein Complex	
Introduction	68
Results.....	71
Figures.....	78
Discussion	88
Materials and Methods	93
Supplemental Figures.....	97
Chapter 3: Tail-anchored Protein Targeting by a Get3 Tetramer	
Introduction.....	102
Results.....	105
Figures	116
Discussion.....	129
Materials and Methods	136
Acknowledgements.....	141
Supplemental Figures.....	142
Conclusion	159
Bibliography.....	163
Appendix A: Fully Automated, Sequential Tilt-series Acquisition with <i>Leginon</i>	181

INTRODUCTION

Protein localization is a process essential to the complexity of life. One of the most fundamental types of localization is integration of proteins into a membrane. Integral membrane proteins perform an essential and diverse set of functions in the cell including transport, signaling and metabolism in all three domains of life. Understanding membrane protein localization in a broader context is fundamental to biology.

By definition, integral membrane proteins must reside in a membrane and generally do so by one or more hydrophobic regions that are thermodynamically stable in a lipid bilayer. For localization this common feature must be considered in how integral membrane proteins get from the point of synthesis by the ribosome to their destination in a membrane through the polar environment of the cytosol. The architecture of integral membrane proteins is as varied as their function with different arrangements of transmembrane segments as well as structures of soluble domains on either side of the membrane. Accommodations must be made for the range of membrane protein configurations during their localization and integration. Membrane localization becomes further complicated in eukaryotic organisms where there are distinct membranes from different organelles in addition to the plasma membrane.

The signal recognition particle (SRP) pathway is a universally conserved mechanism for targeting not only proteins integrated into the membrane but also secreted proteins (reviewed Shan and Walter, 2005). Proteins targeted by the SRP pathway contain an amino-terminal (N-terminal) hydrophobic signal sequence, and as a nascent protein is synthesized by the ribosome the SRP binds the sequence. The complex of SRP and the ribosome nascent chain (RNC) binds to the SRP receptor (SR), at which point the RNC is delivered to the Sec translocon for

integration. Many integral membrane proteins are localized by the SRP pathway, but there are important exceptions.

TA Protein Translocation

In 1993 a review by Kutay et al. recognized that a class of proteins with a transmembrane domain (TMD) at their extreme carboxyl-terminus (C-terminus) would be inaccessible to the SRP pathway. These tail-anchored (TA) proteins lack an N-terminal signal sequence so only the TMD could act as a signal anchor. The proximity of the signal anchor to the C-terminus causes it to be occluded by the ribosome during synthesis and, therefore, unavailable to the SRP during translation. Thus TA proteins must be targeted post-translationally to the membrane for integration. TA proteins are an important class of proteins comprising 2–3% of open reading frames in humans and 1% in yeast and carry out a wide array of functions including vesicle fusion, apoptosis regulation and enzymatic activity (Beilharz et al., 2003; Kalbfleisch et al., 2007).

The model system for TA protein translocation has been the membrane bound form of the heme-binding electron transfer protein cytochrome *b₅*. The first mechanism shown for membrane integration of a TA protein was the spontaneous insertion of cytochrome *b₅* in vitro into liposomes and other membranes (Dailey and Strittmatter, 1978; Takagaki et al., 1983). In contrast to most TA proteins, the TA of cytochrome *b₅* is only moderately hydrophobic and cytochrome *b₅* can exist in a soluble form (Brambillasca et al., 2006). When it was shown that the tail-anchored synaptobrevin, unlike cytochrome *b₅*, requires ATP and a proteinaceous factor at the membrane for insertion in vitro, it became clear that there are other mechanisms for TA insertion (Kutay et al., 1995). The behavior of cytochrome *b₅* and similar

TA proteins is likely not the primary route for TA protein insertion (Mandon and Gilmore, 2007).

Work on finding a general mechanism for TA protein translocation continued with a number of important criteria for their targeting being defined. SRP was shown to not be necessary for proper targeting (Kutay et al., 1995). Furthermore, TA protein targeting to the ER was shown to be independent of the Sec components in both their co- and post-translational translocating forms (Kutay et al., 1995; Steel et al., 2002; Yabal et al., 2003) although the exact nature of this property continued to be debated (Abell et al., 2004). More than a decade after Kutay et al. published their review about TA proteins, the process of their biogenesis was still enigmatic.

A Targeting Factor for TA Proteins

In 2007 the first component of the pathway currently considered to be the mechanism of TA protein translocation was discovered. Stefanovic and Hegde (2007) and Favaloro et al. (2008) demonstrated that a 40 kDa putative ATPase specifically bound TA-proteins and facilitated their translocation in an energy and membrane-associated protein dependent manner, meeting the criteria found for insertion of synaptobrevin. Stefanovic and Hegde (2007) named the complex associated with TA proteins the Transmembrane domain Recognition Complex (TRC) and the targeting factor TRC40.

A large body of work on the gene encoding TRC40 preceded its discovery as the TA protein targeting factor. The first identified gene encoding a TRC40 homolog was annotated in 1992 as a homolog of ArsA, a bacterial ATPase arsenite transport protein, as part of the *Caenorhabditis elegans* genome sequencing project (Sulston et al., 1992). Kaur and Rosen (1992) noted this as the only known eukaryotic homolog of ArsA, and Koonin (1993) recognized the *C. elegans* gene was different from the bacterial *Escherichia coli* and *Mycobacterium leprae* ArsA

genes in that it encoded only one, not two putative ATPase domains. Another eukaryotic homolog of ArsA was identified in *Saccharomyces cerevisiae* as part of the sequencing of chromosome IV, revealing close homology to the *C. elegans* sequence (Boskovic et al., 1996). Kurdi-Haidar et al. (1996) isolated the *Homo sapiens* homolog of ArsA and showed it was universally expressed in different tissues.

In *S. cerevisiae* deletion of the open reading frame *YDL100c*, encoding the ArsA homolog, showed a growth defect phenotype in the presence of metals (zinc, copper and cobalt) and elevated temperatures (Zúñiga et al., 1999). Zúñiga et al. (1999) suggested *YDL100c* was involved in export of toxic compounds, similar to the function of *E. coli* ArsA, and was upregulated as part of the heat shock response. Shen et al. (2003) added to the idea *YDL100c* was involved in metal and heat tolerance including arsenite, chromium and vanadate and annotated *YDL100c* as *ARR4*.

Schuldiner et al. (2005) looked at *S. cerevisiae* epistatic genetic interactions through growth defects in double knockouts of different genes and focused on Arr4 as an example for identification of components in novel pathways. The study implicated Arr4 in a Golgi to ER retrieval (GET) pathway and changed the name Arr4 to Get3. They also identified two other components of the pathway that together formed a receptor in the ER membrane, which they called Get1 and Get2. A subsequent study showed *get3* mutants suppressed a phenotype of a temperature sensitive *npl4* mutant, a gene involved in the ER-associated degradation of proteins by the ubiquitin-proteasome system, and $\Delta get3$ suppresses $\Delta get1$ and $\Delta get2$ phenotypes (Auld et al., 2006). Furthermore, Get3 was shown to associate with the ER membrane through Get1 and Get2 (Auld et al., 2006). Get3 was demonstrated to bind a chloride-transport protein Gef1 and could regulate copper levels, a biochemical link to the

metal related growth defect phenotypes for $\Delta get3$ (Metz et al., 2006). Get3 was also shown to function as a guanine-triphosphate (GTP) exchange factor (GEF) like G protein-coupled receptors, directly binding a G protein Gpa1 in a nucleotide dependent manner (Lee and Dolhman, 2008).

Studies from other organisms examined metal tolerance as well as additional roles for the eukaryotic ArsA homolog. The homolog was shown to be essential in *Mus musculus* with a knockout being lethal at the embryonic stage of development (Mukhopadhyaya et al., 2006). Work on the *C. elegans* homolog by Tseng et al. (2007) focused on the idea that the homolog was involved in metal tolerance, similar to bacterial ArsA and previous studies of the *S. cerevisiae* homolog, by purifying the *C. elegans* homolog and showing in vitro ATPase activity was stimulated in the presence of arsenite. The *C. elegans* knockout of the ArsA homolog reversibly went into dauer, arresting development in the L1 stage due to a defect in regulation of insulin secretion (Kao et al., 2007). Human melanoma cells were sensitive to arsenite and cisplatinin, a cancer therapeutic, when Get3 was downregulated by siRNA (Hemmingsson et al., 2008).

When Stefanovic and Hegde (2007) and Favaloro et al. (2008) identified the mammalian ArsA homolog (TRC40) to be a factor for TA protein targeting, Schuldiner et al. (2008) revisited their previous work with the *S. cerevisiae* homolog (Get3) (Schuldiner et al., 2005). They confirmed the GET pathway as a TA protein targeting pathway and modified the basis of the acronym to Guided Entry of Tail-anchored proteins. The proteins Get1/2 were then shown to form the membrane protein receptor for Get3. In $\Delta get1/2$ strains of *S. cerevisiae*, cytosolic aggregates of TA proteins bound to Get3 and failed to localize TA proteins to the ER membrane in vivo. Insertion of TA proteins by Get3 into microsomes was also dependent on Get1/2 in vitro. Further studies by Jonikas et al. (2009) in *S. cerevisiae* showed strong

interactions between Get3 and the proteins Yor164 and Mdy2, which were then named Get4 and Get5. The $\Delta get4$ and $\Delta get5$ phenotypes were masked by $\Delta get3$, and $\Delta get4$ and $\Delta get5$ strains mislocalized ER destined TA proteins to mitochondria. Extracts used for in vitro translocation assays from $\Delta get5$ strains failed to insert TA proteins into microsomes. Get3 was shown to interact with Get4/5 by immunoprecipitation. This evidence taken together suggested Get4/5 worked upstream of Get3 with Get3 at the center of a pathway chaperoning TA proteins from Get4/5 to the membrane receptor Get1/2. The many other functions attributed to TRC40/Get3 have yet to be explained as either directly or indirectly related to TA protein targeting or as alternative functions, and Get3 is now generally accepted as a targeting factor for TA proteins to the ER.

Get3 as a Nucleotide Hydrolase

The close sequence homology between Get3 and ArsA is confined mostly to their nucleotide hydrolase domains (NHD) even though they were originally annotated as homologous and having similar functions. Proteins that bind and hydrolyze nucleoside-triphosphate (NTP) are ubiquitous and involved in a multitude of processes essential to life. NTP binding and hydrolysis is employed by a wide range of proteins with a variety of localizations, biological functions and molecular mechanisms. NTPases include chaperones and targeting factors with similar function to Get3 such as heat shock proteins and the SRP, respectively. The energy of NTP hydrolysis is often used to drive conformational changes that regulate interactions with other proteins, and the structure of NTPases have revealed how their molecular mechanism is related their function. Structural information about Get3 could serve as the basis for understanding how nucleotide hydrolysis is coupled to binding TA proteins and the process by which they are targeted for membrane integration at the ER.

NTPases can be classified by their structure and divided into a number of distinct folds including the P-loop (phosphate-binding loop) containing NTPase fold (Saraste et al., 1990). There are a number of distinct sequence and structural groups within P-loop containing NTPases, and the NHD of ArsA, the closest sequence homolog to Get3, falls within the superclass of P-loop containing GTPases (Bourne et al., 1991; Koonin, 1993). The P-loop containing GTPase superclass also includes ATPases, contrary to their name, but will be still be referred to as P-loop GTPases to avoid overlapping nomenclature with the larger group of P-loop NTPases.

The structure of the NHD of P-loop containing NTPases is a mix of mostly alternating α -helices and β -strands with conserved motifs residing in the loops that connect them. The α/β fold, which describes the mixed secondary structure, forms a tertiary structure that sandwiches a seven strand β -sheet between two groups of α -helices exposing the connecting loops on the edges. Motifs within the loops are important for binding and hydrolysis of nucleotide as well as sensing nucleotide state and are located on one edge of the β -sheet. These conserved motifs define the P-loop GTPase superclass from other P-loop NTPases.

Motifs are illustrated in Figure 1 and include the P-loop, also called a Walker A motif or G1 region, with the consensus sequence GxxxxGK(S/T) (x for any amino acid) that forms a loop after strand 1 (β 1). This motif interacts with the phosphates of a bound nucleotide (Walker et al., 1982). In P-loop GTPases the switch II region, also called a Walker B motif or G3 region, following β 4 has the sequence hhhhDxxG (h for hydrophobic amino acids) and is involved in magnesium ion coordination and binding the γ -phosphate of the nucleotide (Walker et al., 1982; Bourne et al., 1991). Conserved features with less consensus in terms of sequence include the switch I or G2 region after β 2 that coordinates a magnesium ion bound to the β -

and γ -phosphates of the nucleotide. Also, a nucleotide specificity motif (N/T)KxD (G4 region) following $\beta 6$ that is conserved for GTPases (Bourne et al., 1991) but with notable exceptions including ATPases like Get3. Main chain interactions from the A-loop (as named in ATPases) or G5 region after $\beta 7$ interact with the nucleotide but are not strictly conserved as a motif.

P-loop GTPases can then be divided into two major classes based on their sequence and structure: the TRAFAC (translation factor) class and the SIMIBI (signal recognition particle, MinD, BioD) class (Leipe et al., 2002). The TRAFAC class includes Ras, EF-Tu and ATPases kinesin and myosin and is defined by having $\beta 2$ on the outermost edge of the core β -sheet. A conserved threonine or serine in the switch I region is followed by an anti-parallel $\beta 3$. Get3 falls within the SIMIBI class of GTPases, which have parallel $\beta 2$ and $\beta 3$ with their positions swapped compared to TRAFAC GTPases. The SIMIBI class is divided into three classifications reflected in the acronym: SRP family, MinD superfamily and BioD family.

Leipe et al. (2002) provide an evolutionary classification of P-loop GTPases based on sequence, which can be used to further define the most appropriate proteins to relate Get3 with molecular mechanism. The MinD superfamily has only the aspartate of (N/T)KxD guanine nucleotide specificity motif conserved and is divided into the MinD/Mrp and Etk families. The MinD/Mrp family has a characteristic P-loop/Walker A motif named the “deviant” Walker A motif, GKGGhGK(S/T), with a lysine in the second position followed by two glycines. The MinD/Mrp family is finally divided into subfamilies of essentially homologous proteins. Get3 is technically part of the ArsA subfamily, but the NifH subfamily and other MinD/Mrp family members are suitable for speculating on how Get3 functions as a nucleotide hydrolase.

The Structure and Mechanism of Related P-loop GTPases

There have been a number of structural and functional studies of members of the MinD/Mrp family. The nitrogenase iron protein (Fe-protein) was the first family member to be structurally characterized (Georgiadis et al., 1992). The Fe-protein fixes nitrogen through conversion of nitrogen gas to ammonia in an ATP hydrolysis dependent electron transfer to a partner molybdenum-iron protein (MoFe-protein). Transfer by the complex occurs through an iron-sulfur cluster in the Fe-protein to the P-cluster and then iron-molybdenum cofactor in the MoFe-protein complex (Howard and Rees, 1994). The structures of the nitrogenase complex between the Fe- and MoFe-proteins in different nucleotide states illustrates the mechanism of the Fe-protein during nucleotide binding and hydrolysis (Schindelin et al., 1997; Tezcan et al., 2005). ATP binding by the Fe-protein is coupled to the formation of the nitrogenase complex by conformational changes that occur when binding the MoFe-protein. Rotation between the two subunits of the Fe-protein form a compact overall structure (Figure 2A & B), presenting a specific binding interface for the MoFe-protein (Schindelin et al., 1997). Similar to the Fe-protein, Get3 interactions with other proteins including TA proteins, Get1/2 and Get4/5 could be regulated by different nucleotide dependent conformations. The key features and conformational changes in the structures of the Fe-protein serve as a basis for examination of Get3 and the MinD/Mrp family as a whole.

Structural studies of the arsenite transport protein ArsA (Zhou et al., 2000; 2001), the closest sequence homolog to Get3, followed the initial structural characterization of the Fe-protein. ArsA binds to an integral membrane protein channel, ArsB, to form an arsenite export mechanism with ArsA as the catalytic subunit. The hypothesis that the tandem repeat of ATPase domains in ArsA could serve to form a pseudo-dimer, fulfilling the homodimerization requirement for the Fe-protein, (Li and Rosen, 2000) was confirmed by the

structure. Arsenite or antimonite is coordinated by conserved histidines and cysteines in the regions connected to the NHD (Zhou et al., 2000). Structures of ArsA in different nucleotide states show significantly less conformational change than the Fe-protein, with minor changes in the switch I and II, P-loop and A-loop regions (Figure 3) (Zhou et al., 2001). This suggests the mechanism of ArsA requires only small alterations in the coordination geometry of arsenite to drive arsenite transport, or the constraints of packing in the crystal used for structure determination preclude large conformational changes (Zhou et al., 2001; Lutkenhaus and Sundaramoorthy, 2003). Additionally, binding of ATP, the ATP analog AMPPNP or the transition state analog ADP·AlF₃ only takes place in one of the two NHDs (Zhou et al., 2001). The model for the mechanism of arsenite transport based on the structures was a reciprocating pump driven by ArsA in one NHD at a time (Zhou et al., 2001). This contrasted to the binding, simultaneous hydrolysis and release mechanism of nitrogenase. As the closest sequence homolog to Get3, the structure and mechanism of ArsA is logical for comparison to Get3.

MinD was the next family member structurally characterized and is part of a regulatory system for septum formation in cell division (Cordell and Löwe, 2001; Hayashi et al., 2001; Sakai et al., 2001). Homodimeric MinD bound to ATP associates with the membrane through a C-terminal amphipathic helix. MinC binds to MinD to inhibit the cell division protein FtsZ but is displaced by MinE, resulting in ATP hydrolysis and dissociation from the membrane by MinD (Lutkenhaus, 2007). MinD exists as a monomer in solution, and structures determined in a variety of nucleotide states were all monomeric (Cordell and Löwe, 2001; Hayashi et al., 2001; Sakai et al., 2001). The mechanism of MinD through ATP hydrolysis was framed in the context of the homodimeric Fe-protein by structural alignment of the monomers to the Fe-protein (Lutkenhaus and Sundaramoorthy, 2003). There is little difference between the ATP

analog AMPPCP and ADP bound states suggesting the AMPPCP bound structure does not completely represent the ATP bound state (Figure 4) (Lutkenhaus and Sundaramoorthy, 2003). From these structures it is difficult to determine what rearrangements take place during nucleotide hydrolysis that regulate membrane binding and association with MinC and MinE. Nevertheless, nucleotide binding features and association with the membrane are topics relevant to Get3.

Structures of another family member, Soj, were determined in apo, ADP and ATP bound states (Leonard et al., 2005). The protein Soj has many parallels in function and structure to MinD including spatial oscillations and polymerization in the cell and close sequence homology, except for the membrane associating C-terminal amphipathic helix of MinD (Lutkenhaus and Sundaramoorthy, 2003). Soj functions as part of a bacterial chromosome partitioning system forming filaments on DNA as a homodimer bound to ATP (Ebersbach and Gerdes, 2005; Ghosh et al., 2006). Hydrolysis and release from DNA is stimulated by the N-terminus of Spo0J, a protein that binds to a specific site in the chromosome. As for MinD, structures of Soj in apo and ADP bound forms were monomeric, but a dimeric ATP bound structure was solved using a hydrolysis deficient mutant (Leonard et al., 2005).

The MinD/Mrp family member ParA is also part of a plasmid partitioning system and is very similar to Soj with a partner protein ParB analogous to Spo0J (Ebersbach and Gerdes, 2005; Ghosh et al., 2006). Structures of ParA (type Ib) in an ATP γ S bound state (Pratto et al., 2008) as well as apo and ADP bound states of a ParA with a specific DNA binding winged helix-turn-helix motif (type Ia) (Dunham et al., 2009) have been determined. Unlike Soj, ParA was dimeric in solution even in apo or ADP bound states. Despite Soj and MinD having similar structure to the monomeric subunit of ParA, the ParA ATP γ S bound dimer structure differs

from the hydrolysis inactive dimer structure of Soj with a more open conformation exposing the nucleotides (Pratto et al., 2008). The ParA apo and ADP structures adopt a range of dimer conformations in different crystal forms mediated through an N-terminal α -helical extension (Figure 6) (Dunham et al., 2009). Although closely related to Soj, the structures of ParA show significant differences in oligomerization related to nucleotide states and in conformation between the subunits of the dimer.

Homodimerization at the NHDs seems to be a prerequisite for ATP hydrolysis in deviant Walker A motif containing proteins, including Get3. The intersubunit interaction by the first conserved lysine in the motif to the nucleotide phosphates is thought to stabilize negative charge in the transition state and be necessary for nucleotide hydrolysis (Figure 2A) (Koonin, 1993; Schindelin et al., 1997). Fe-protein dimerization is covalently linked by the iron-sulfur cluster, but additional hydrophobic and hydrogen-bonding interactions between the subunits are sufficient to maintain the dimeric state (Howard et al., 1989; Georgiadis et al., 1992). Covalent dimerization is also present in ArsA, and similar to the Fe-protein additional interactions are present (Zhou et al., 2000). The ParA dimer lacks the covalent linkage of the Fe-protein and ArsA and is stabilized by interactions through hydrophobic surfaces and hydrogen-bonding between several residues (Pratto et al., 2008; Dunham et al., 2009). MinD and Soj function with a different mechanism where they go from a monomeric to a dimeric state upon binding ATP (Hu and Lutkenhaus, 2003; Leonard et al., 2005). Overall, Mrp/MinD family members related to Get3 function as homodimers, and there are several models for the nature of dimerization depending on the protein.

The deviant Walker A motif not only binds nucleotide through α and β -phosphates but is involved in catalysis (Koonin, 1993). The conformation of the P-loop when nucleotide bound

is generally invariant, and the Fe-protein, ArsA and MinD structures show little change in P-loop position between nucleotide states (panels C & D from Figures 2 & 3, Figure 4) (Schlessman et al., 1998; Zhou et al., 2001). There are differences in P-loop conformations between apo and nucleotide bound states as seen in Soj (Figure 5), and ParA structures where extended or disordered forms of the P-loop move to form close interactions with the α and β -phosphates of ADP (Figure 6C & D) (Leonard et al., 2005; Pratto et al., 2008; Dunham et al., 2009). In the structure of the Fe-protein bound with the transition state analog ADP·AlF₄⁻, the first conserved lysine of the motif interacts with β and γ -phosphates of the bound nucleotide in the opposing subunit, an interaction not seen in the ADP bound structure (Figure 7A) (Schindelin et al., 1997). The Fe-protein structures are the standard for examining MinD/Mrp family members. Comparison of dimeric ATP bound Soj to the transition state of the Fe-protein also showed the deviant Walker A lysines in position to interact with the α and β -phosphates of the ATP molecule (Figure 7B) (Leonard et al., 2005). The interaction is not present in ParA dimer structures (Pratto et al., 2008; Dunham et al., 2009), but mutation of the lysine to alanine shows reduced ATPase activity (Pratto et al., 2008). Biochemical analysis of MinD has also shown the deviant Walker lysine essential for catalysis (de Boer et al., 1991), but due to the monomeric states of the structures the interaction is not seen (Cordell and Löwe, 2001; Hayashi et al., 2001; Sakai et al., 2001). The structures of ArsA with ADP·AlF₃, ATP or AMPPNP bound lack this interaction (Zhou et al., 2001). P-loop interactions, especially the deviant Walker A lysine, are crucial for interpreting the structures of MinD/Mrp family members.

The switch II region senses the γ -phosphate of bound ATP to drive conformational changes from the apo or ADP bound state and interacts with the nucleotide phosphates and

magnesium ion. In structures of the Fe-protein, the differences in switch II interactions between the apo and nucleotide bound forms result in conformational changes to proximal parts of the protein that govern function (Figure 2 & 8A). The rearrangement of the switch II region serves as a mechanism for transmitting nucleotide state and positions the iron-sulfur cluster closer to the MoFe-protein for electron transfer (Schindelin et al., 1997; Schlessman et al., 1998). Additionally, an aspartate from the opposite switch II region could position a water for a catalytic attack on the nucleotide. In *ArsA*, arsenite or antimonite is coordinated by histidines several residues C-terminal to the switch II region allowing for direct movement of these residues (Zhou et al., 2000). There is little difference in the structures of *ArsA* in different nucleotide states so switch II movements are not seen (Figure 3 & 8B) (Zhou et al., 2001; Lutkenhaus and Sundaramoorthy, 2003). The ADP and ATP bound structures of MinD also show little conformational change suggesting either the difference in nucleotide is not related to the function of the protein or there is crystallographic constraint of the conformation (Hayashi et al., 2001). Comparison of the *Soj* ADP and ATP structures shows little difference in switch II as well, and with ATP bound the P-loop is shifted down and γ -phosphate moved out suggesting this does not reflect the actual ATP bound state (Figure 5) (Leonard et al., 2005). *ParA* switch II regions have differences in apo and nucleotide bound structures but not ATP γ S and ADP structures (Figure 6C & D) (Pratto et al., 2008; Dunham et al., 2009). Although suggested for other proteins, only the structure of the Fe-protein shows the conformational change in the switch II region in response to presence of the γ -phosphate making it the best model for this action.

Switch I is similar to switch II in terms of interactions with nucleotide phosphates and magnesium and in conformational changes dependent on nucleotide state. In the MinD/Mrp family the switch I loop contains a conserved aspartate, present in *Get3*, that appears to

position water for nucleophilic attack on the γ -phosphate unlike other ATPases where a switch II residue performs this function (Leipe et al., 2002). The structures of MinD with AMPPCP bound and ParA with ATP γ S bound illustrate this interaction (Figure 9) (Hayashi et al., 2001; Pratto et al., 2008). In the Fe-protein, the conformation of switch I shows significant variability in different nucleotide states (Schlessman et al., 1998). Although a switch II residue could position the nucleophilic water, the conserved switch I aspartate is also present that could perform this function (Schindelin et al., 1997; Leonard et al., 2005). The hydrolysis inactivating mutation in the Soj ATP bound structure was to the conserved switch I aspartate, and the aspartate was suggested to position a nucleophile water (Leonard et al., 2005). In the structures of ArsA the interaction is not present. Analogous to switch II, changes in switch I interactions result in conformational changes, and of additional importance is a conserved aspartate involved in nucleotide hydrolysis.

Significant conformational changes in the relative position of the subunits in the homodimeric Fe-protein or changes in oligomeric state of MinD and Soj based on nucleotide state are critical for their function. The Fe-protein undergoes a $\sim 13^\circ$ rotation between the two subunits in the transition state to form the interaction site for the MoFe-protein (Figure 2A & B) (Schindelin et al., 1997). The interaction of the dimeric Soj with Spo0J upon ATP binding is analogous to the Fe-protein and MoFe-proteins in the nitrogenase complex. In the case of Soj, dimerization from monomers rather than a rotation between subunits makes Soj competent for Spo0J binding (Leonard et al., 2005). For MinD, ATP binding and dimerization have been proposed to expose the C-terminal amphipathic helix for membrane association (Hu and Lutkenhaus, 2003). Finally, the many conformations of ParA demonstrate the flexibility of the dimer in apo or ADP bound states (Figure 6A & B) (Dunham et al., 2009).

The different models for changes in the homodimeric state are key for understanding the function of Get3.

Through sequence homology the MinD/Mrp family of P-loop GTPases, including Get3, appears to be closely related by evolution (Leipe et al., 2002). Examination of previous structural studies serves as a basis for examining the structure and molecular mechanism of Get3. Both the common themes and appreciable differences in the structure and function of related GTPases is important for connecting the structure of Get3 to its function as a nucleotide hydrolase. This defines its mechanism and part of the role it plays in TA protein translocation.

Binding TA Proteins

How TA proteins interact with Get3 is a fundamental question key to understanding the GET pathway. The hydrophobic TA presumably is involved in an interaction with a hydrophobic surface presented by Get3 and is also protected from the aqueous environment of the cytosol. Furthermore, nucleotide state and hydrolysis would be a mechanism for altering the interactions that bind and then release the TA protein. Although Stefanovic and Hegde proposed a putative TA protein binding site in Get3 based on a hydrophobic patch of residues (Stefanovic and Hegde, 2007), there are no obvious motifs for protein interactions in the sequence of Get3. The structure of Get3 is an obvious platform for identification and characterization of the TA protein interaction site.

In terms of function the SRP is the closest molecular assembly to Get3. SRP54/Ffh is the universally conserved protein component of the SRP that binds to the hydrophobic signal sequence or anchor (reviewed in Keenan et al., 2001; Doudna and Batey, 2004) similar to Get3 binding the TA. Interestingly, SRP54/Ffh is also a nucleotide hydrolase in the same SIMBI

class as Get3 (Leipe et al., 2002). Unlike homodimeric MinD/Mrp family members, SRP54/Ffh has GTPase activity stimulated by heterodimerization with the SR; however, the GTP hydrolase domains of the SRP and SR are homologous in sequence and structure. Three domains comprise SRP54/Ffh: an N-terminal four-helix bundle composing the N domain, the GTP hydrolyzing G domain and the signal sequence binding methionine-rich M domain.

Structures of SRP54/Ffh have given insight into the mechanism of signal sequence binding (Figure 10) (reviewed in Keenan et al., 2001; Doudna and Batey, 2004). The N and G domains are in close contact, but only a flexible linker connects them to the M domain. The M domain forms a hydrophobic pocket within a group of α -helices in conjunction with a flexible “finger loop” that could accommodate the signal sequence. How the NG domain couples nucleotide state with M domain signal sequence binding is a complex question given the many interactions of the three domains with the SRP RNA, ribosome, SR and the translocon. The SRP serves as an example of how a compact mechanism for protecting a diverse set of hydrophobic sequences from the cytosol can be made from a flexible hydrophobic pocket.

Summary

Protein translocation is a fundamental process in every organism. The SRP pathway is a universal mechanism for translocating many, but not all integral membrane proteins. TA proteins are a large and diverse class of integral membrane proteins that cannot be co-translationally translocated and are instead post-translationally targeted via the GET pathway. The central component of the GET pathway is the MinD/Mrp family member ATPase Get3, which functions between Get4/5 and the ER membrane receptor Get1/2. Targeting by Get3 is dependent on binding the TMD of TA proteins and ATP hydrolysis. This work gives insights into TA protein targeting by Get3 based on structural studies. Chapter 1 presents the

structure of Get3 and derives a model for tail-anchored protein binding coupled to nucleotide hydrolysis. Chapter 2 describes expression, purification and crystallization experiments on the complex of Get3 with TA proteins. Chapter 3 presents the structure of a tetrameric archaeal homolog of Get3, evidence archaeal Get3 can bind TA proteins and describes a model where a central hydrophobic chamber of the Get3 tetramer sequesters TA proteins.

Figures

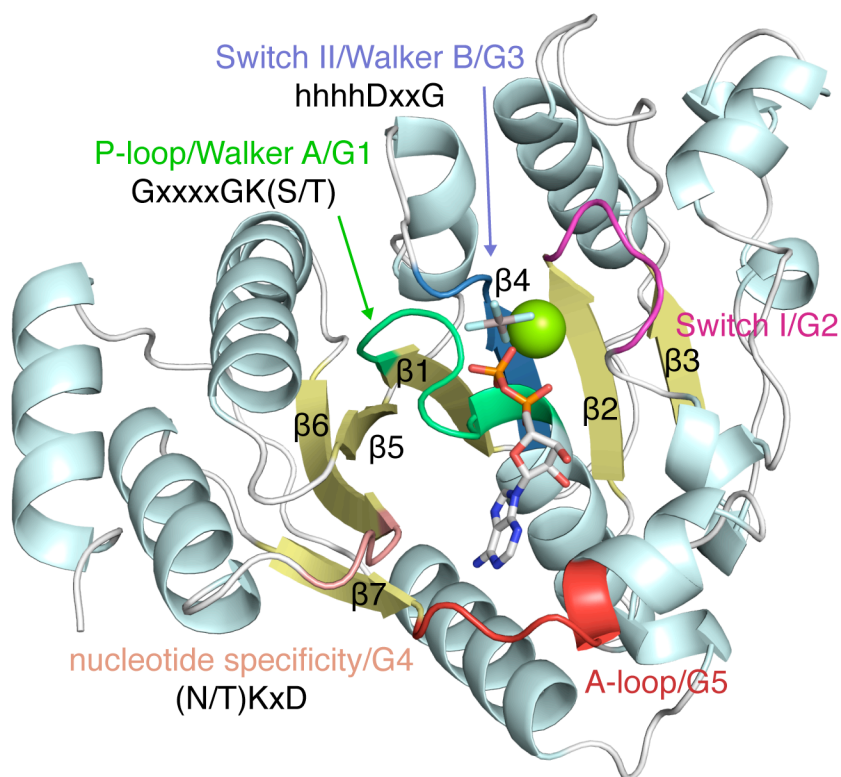


Figure 1. The structure of conserved P-loop GTPase motifs. A cartoon diagram of the monomeric subunit of the Fe-protein (PBDID 1N2C) bound to Mg (green sphere) and the transition state analog ADP·AlF₄⁻ (sticks) is a model for describing the P-loop GTPase motifs. The α -helices are colored cyan, β -strands yellow and loops white. Motifs are colored: P-loop/Walker A/G1 (green), Switch I/G2 (magenta), Switch II/Walker B/G3 (blue), nucleotide specificity/G4 (pink) and A-loop/G5 (red).

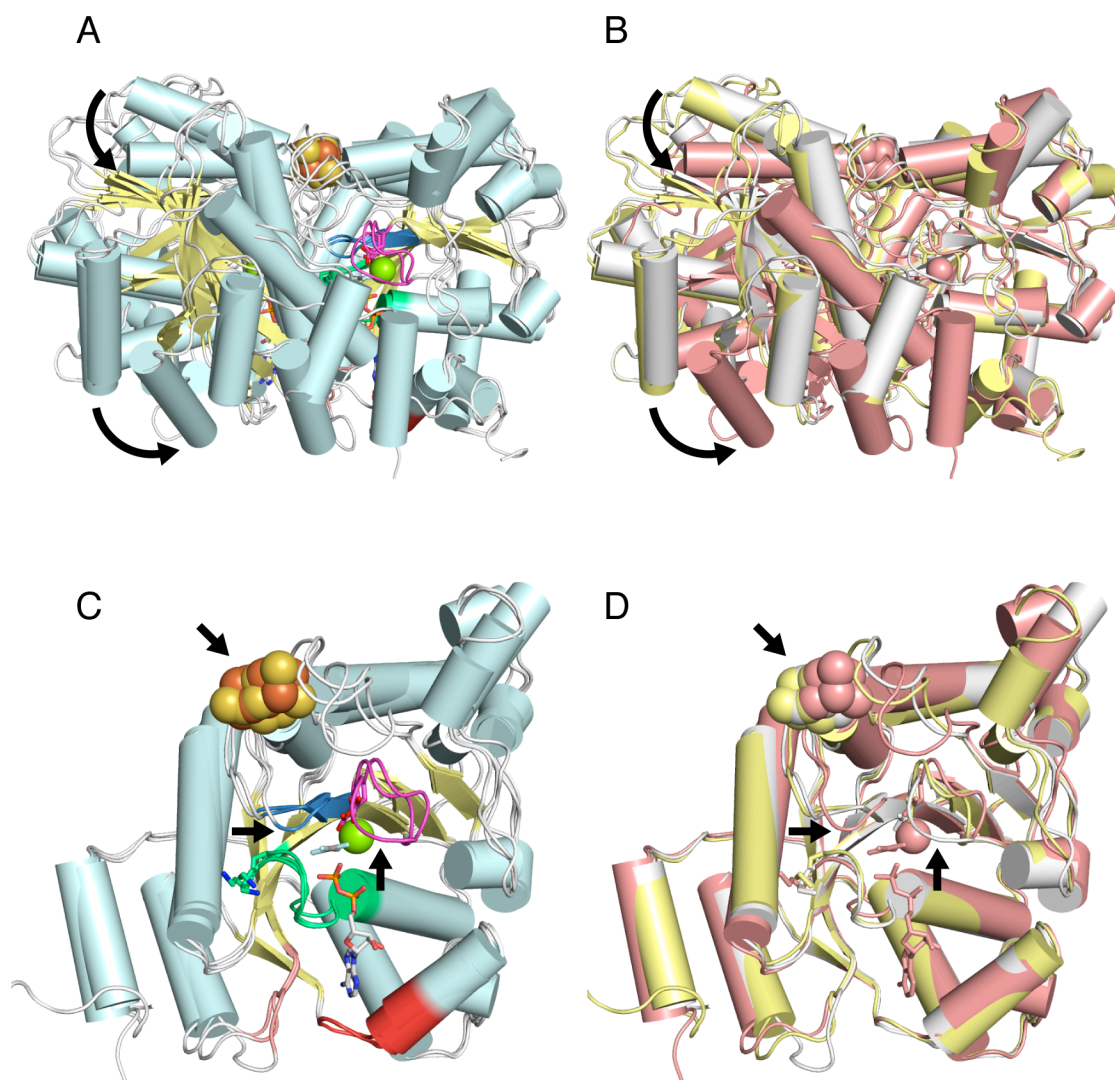


Figure 2. Structural alignment of the Fe-protein in different nucleotide states. (A) Fe-protein dimer structure alignment shown as a cartoon diagram with cylindrical helices, nucleotide in sticks and the spheres for the iron-sulfur cluster (orange and yellow) and magnesium (green). The color scheme for motifs and secondary structure is the same as Figure 1. Apo (PBDID 2NIP, 1CP2) and transition state analog ADP·AlF₄⁻ bound (PDBID 1N2C) structures are aligned by the right subunit to show relative rotation in the left subunit in different nucleotide states (indicated with arrows). (B) Structures from (A) are colored red (2NIP, apo), white (1CP2, apo) and gold (1N2C, ADP·AlF₄⁻). (C) As in (A) with the left subunit cut away to show the NHD. Movements in the switch II region (blue, indicated with arrow) and iron-sulfur cluster (spheres, indicated with arrow) occur in the transition state and switch I flexibility is apparent in apo forms (magenta, indicated with arrow) (D) NHD from (C) colored as in (B).

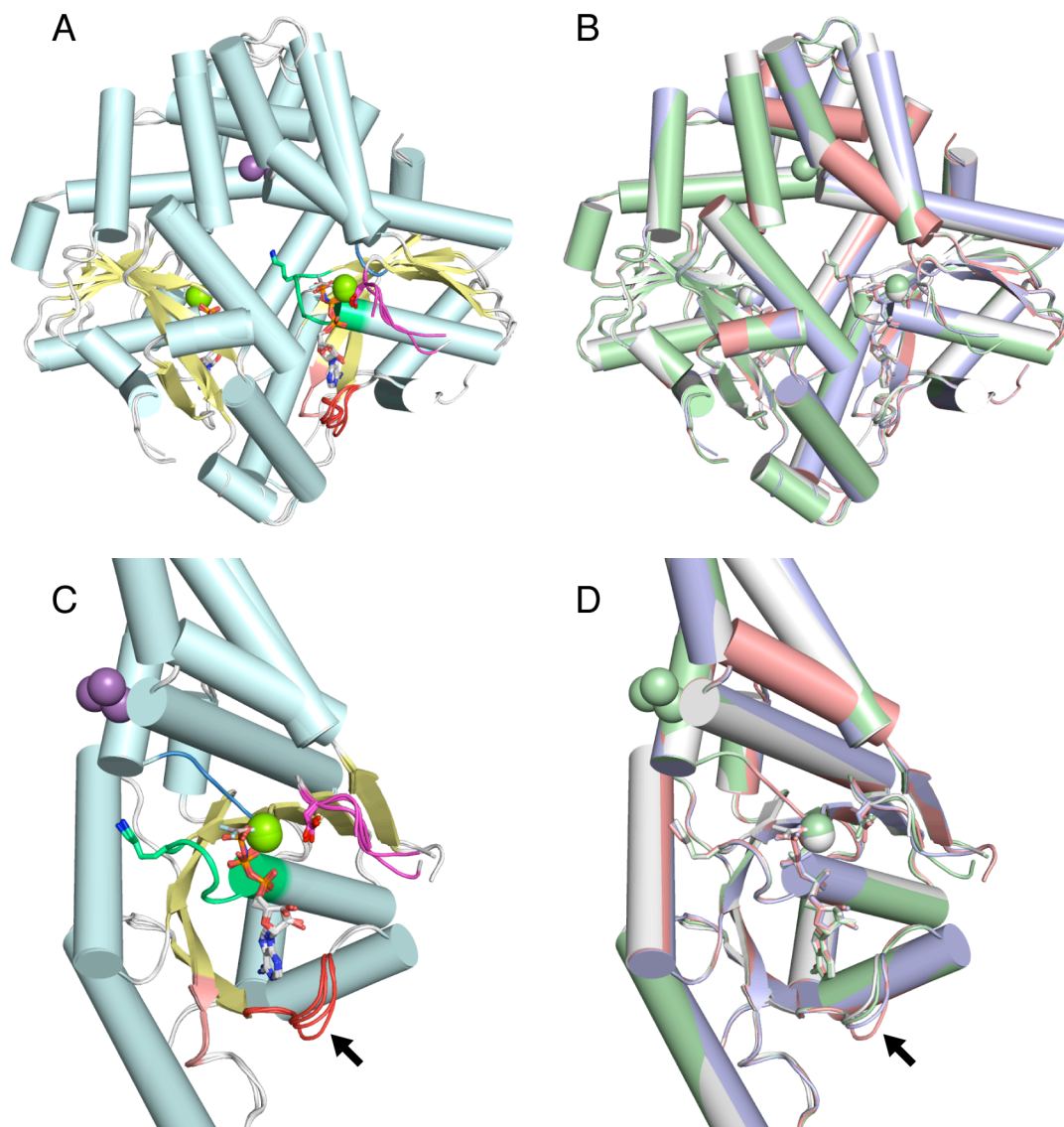


Figure 3. Structural alignment of ArsA in different nucleotide states. (A) ArsA structures represented and aligned as in Figure 2 with antimony shown as purple spheres. ADP (PDBID 1F48), ADP·AlF₃ (PDBID 1IHU), ATP (PDBID 1II0) and AMPPNP (PDBID 1II9) bound structures show little overall conformational differences in between nucleotide states. (B) Structures from (A) are colored green (1F48, ADP), white (1IHU, ADP·AlF₃), red (1II0, ATP) and blue (1II9, AMPPNP). (C) As in (A) with the left subunit cut away to show the NHD. The NHDs are almost identical with minor changes in the position of the A-loop (indicated with arrow) (D) NHD from (C) colored as in (B).

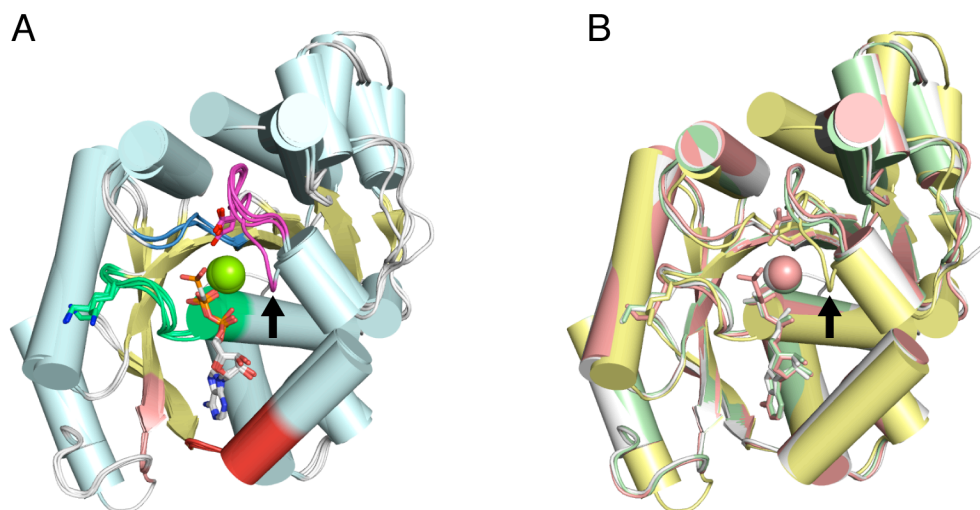


Figure 4. Structural alignment of MinD in different nucleotide states. (A) MinD structures represented as in Figure 2. Structures in apo (PDBID 1HYQ), ADP bound (PDBID 1G3Q, 1ION) and AMPPCP bound (PDBID 1G3R) forms show little differences in conformation except for switch I in apo form (indicated with arrow). (B) Structures from (A) are colored gold (1HYQ, apo), green (1G3Q, ADP), red (1G3R, AMPPCP) and white (1ION, apo).

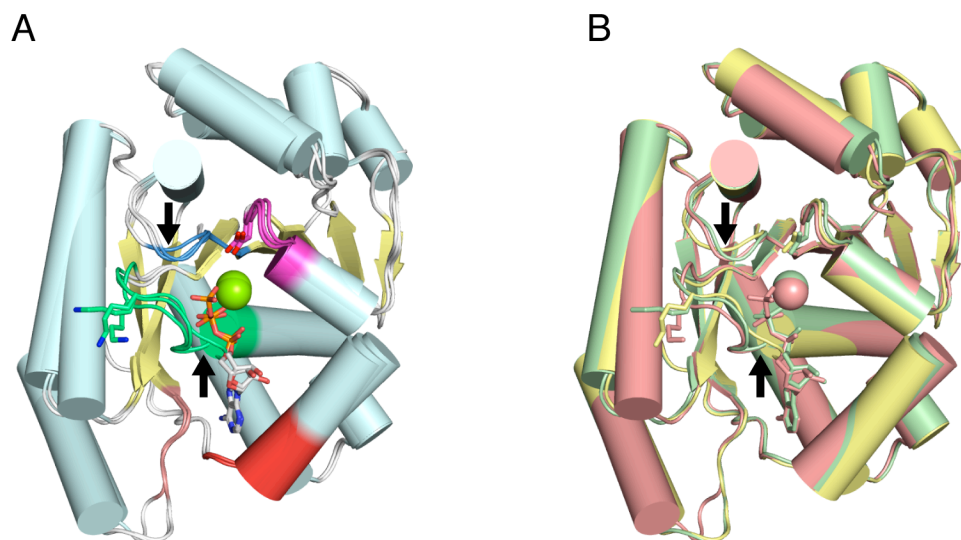
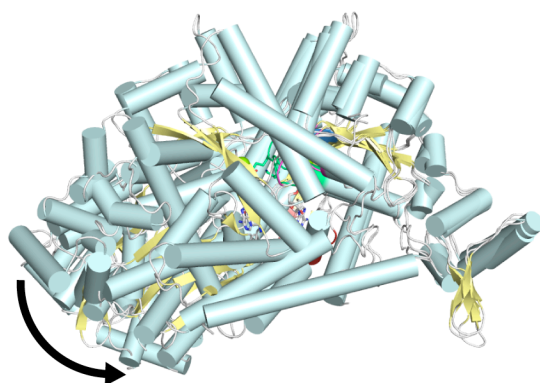
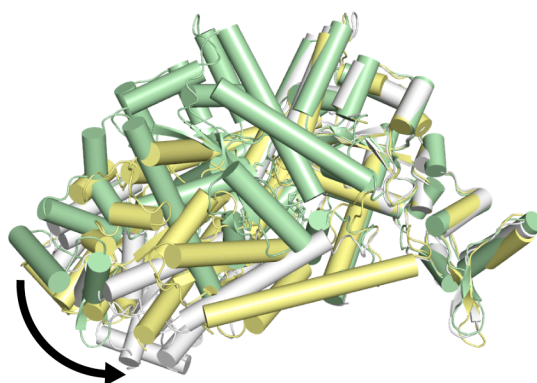


Figure 5. Structural alignment of Soj NHDs in different nucleotide states. (A) Soj structures represented and aligned as in Figure 2. Structures apo (PDBID 1WCV), ADP bound (PDBID 2BEJ) and ATP bound (PDBID 2BEK) forms are shown. The apo form Switch II has minor differences in conformation (blue, indicated with arrow) and P-loop (green, indicated with arrow). (B) Structures from (A) are colored gold (1WCV, apo), green (2BEJ, ADP) and red (2BEK, ATP).

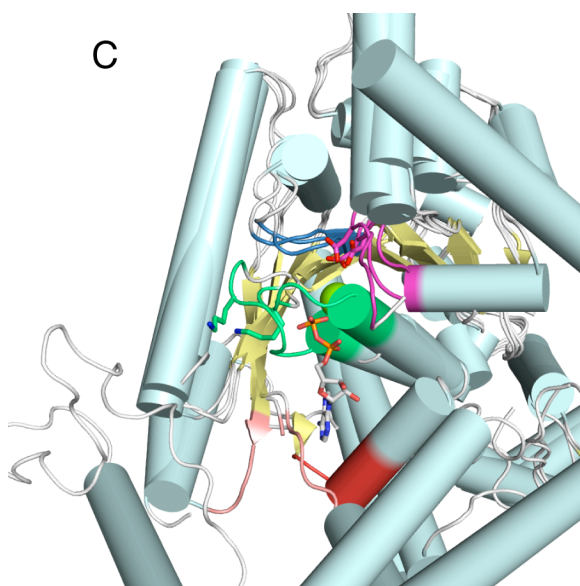
A



B



C



D

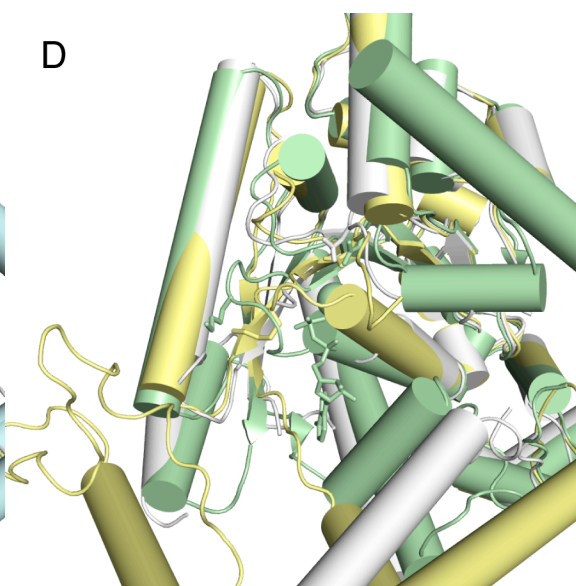


Figure 6. Structural alignment of ParA in different nucleotide states. (A) ParA dimer structures represented and aligned as in Figure 2. Apo (PDBID 3EZ7, 3EZ9 chain B) and ADP bound (PDBID 3EZ2) forms show a range of different conformations for the left subunit relative to the right subunit (indicated by arrow). (B) Structures from (A) are colored gold (3EZ7, apo), white (3EZ9) and green (3EZ2, ADP). (C) As in (A) with the left subunit cut away to show the NHD. The NHDs motifs also show a wide variety of conformations. (D) NHD from (C) colored as in (B).

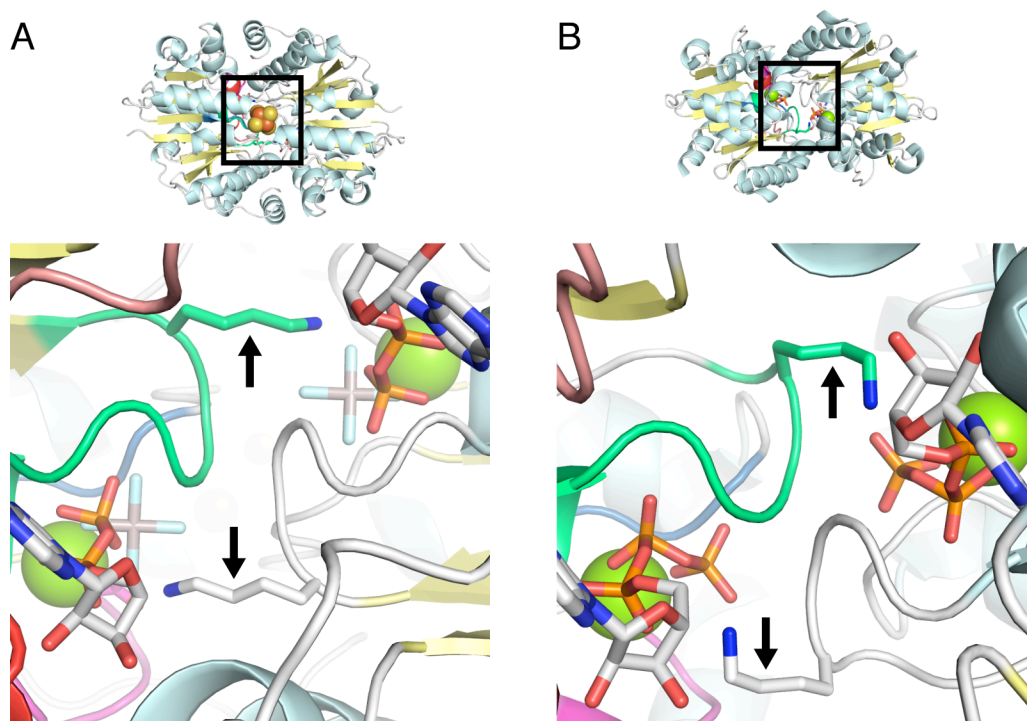


Figure 7. Deviant Walker A motif lysines. (A) Inset (top) shows the top of the Fe-protein dimer relative to Figure 2. Bottom panel is close-up view (box from inset) of the deviant Walker A motif P-loop of the Fe-protein with ADP·AlF₄⁻ bound (PDBID 1N2C). Representation is the same as Figure 1 with the motifs colored in the left subunit. The conserved lysines (indicated with arrows) in the P-loop reach across the dimer interface to the opposite NHD to interact with the nucleotide phosphates. (B) As in (A) for the Soj dimer with ATP bound (PDBID 2BEK).

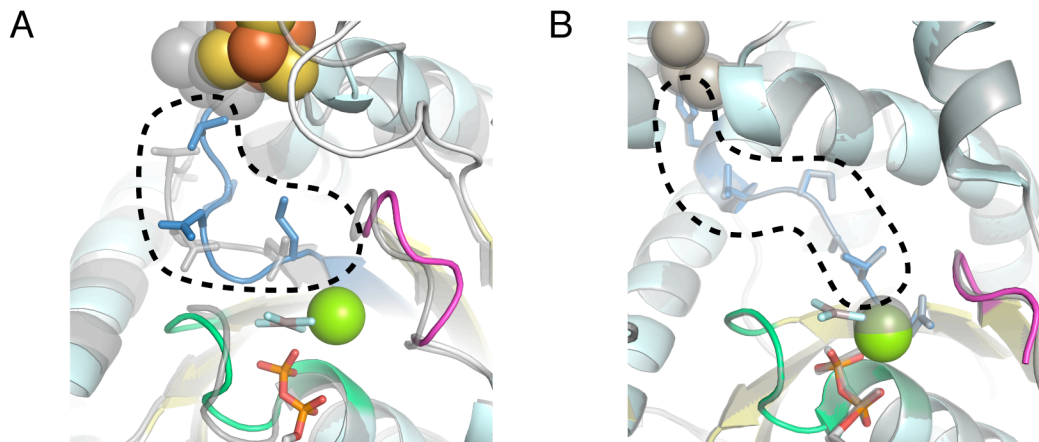


Figure 8. Switch II conformational changes with nucleotide state. (A) The NHD domain of the Fe-protein (PDBID 1N2C) shown as in Figure 1 with a stick representation of switch II side chains (indicated with dotted circle) and iron-sulfur cluster in spheres. In transparent gray is the apo form of the Fe-protein (PDBID 2NIP) showing the differences in conformations depending on nucleotide state. (B) ArsA bound to ADP·AlF₃ (PDBID 1HIU) compared to the ADP bound form (PDBID 1F48) shown as for (A) with metal ligands in spheres. There is little change in the conformation between the two states.

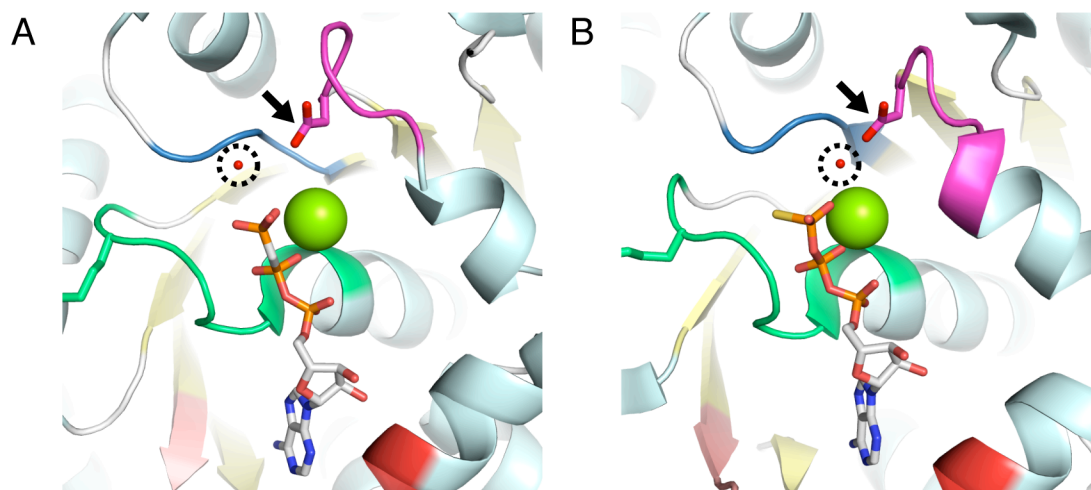


Figure 9. Catalytic water positioned by Switch I aspartate. (A) NHD of MinD with AMPPCP bound (PDBID 1G3R) represented as in Figure 1. The conserved aspartate from switch I (stick representation, indicated with arrow) positioning a water (red dot, indicated with dashed circle) for nucleophilic attack on the γ -phosphate. (B) As in (A) for ParA with ATP γ S bound (PDBID 2OZE).

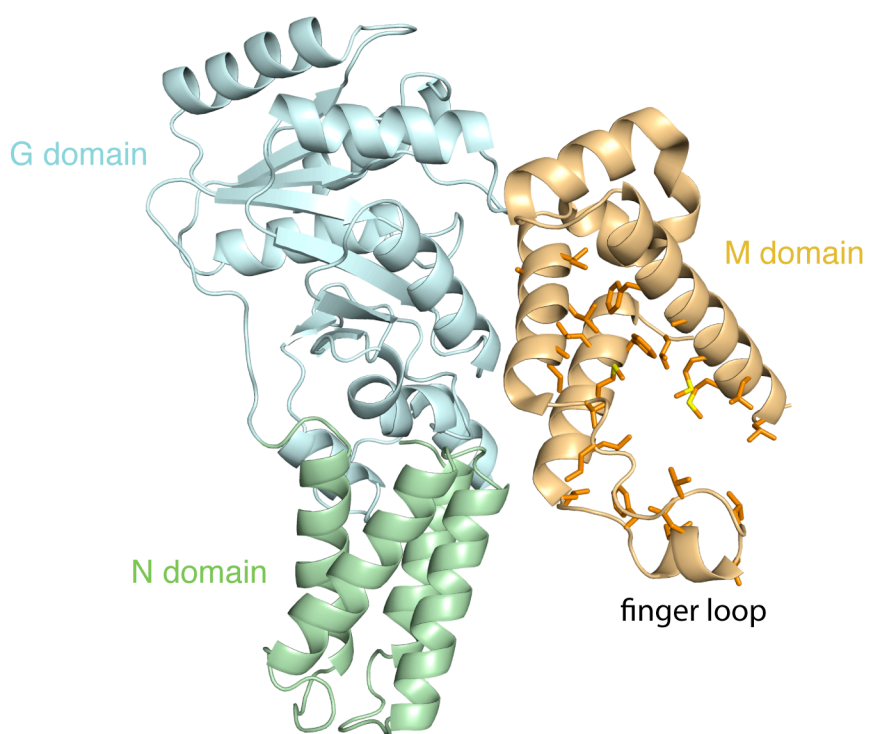


Figure 10. SRP54/Ffh domains and hydrophobic binding pocket. Cartoon diagram of Ffh (PDBID 2FFH) shows the three domains: four-helix bundle N domain (green), GTPase G domain (blue) and methionine-rich M domain (orange). The N and G domains have close interactions compared to the M domain. Hydrophobic residues (stick representation) line a hydrophobic groove formed by the helices and finger loop.

Chapter 1

A MODEL FOR TAIL-ANCHORED PROTEIN BINDING BY GET3

Introduction

Tail-anchored (TA) proteins represent a large and diverse class of integral membrane proteins that are found in all organisms. These include numerous types of proteins, such as SNAREs, apoptosis factors, and protein translocation components. TA proteins are characterized by having a single transmembrane helix (TM) at their extreme C terminus. Due to this topological constraint, these proteins are not able to follow the SRP-dependent co-translational pathway that typifies most integral membrane proteins. Instead, these proteins must find their correct membrane for insertion post-translationally (reviewed in (Kutay et al., 1993) and (Borgese et al., 2007)).

The ATPase Get3 was the first protein identified directly involved in TA targeting and is part of the Get pathway (now known as Guided Entry of Tail-anchored proteins) that also contains the ER membrane proteins Get1/2 and the putative ribosome receptor proteins Get4/5 (Schuldiner et al., 2005; Stefanovic and Hegde, 2007; Favaloro et al., 2008; Schuldiner et al., 2008; Jonikas et al., 2009). Multiple studies have shown that Get3 binds directly to the hydrophobic tail-anchors and, in conjunction with ribosome and endoplasmic reticulum (ER) factors, utilizes an ATP cycle to bind and then release TA proteins at the ER membrane.

Get3 was originally annotated Asna-1/Arr4p due to its apparent homology ($\approx 25\%$ identity) to the bacterial arsenite transporter component ArsA (Boskovic et al., 1996). Get3 homologues had been implicated in a diverse set of functions now presumed to be linked to the correct localizations of TA proteins (Shen et al., 2003; Kao et al., 2007; Tseng et al., 2007;

Hemmingsson et al., 2008). Get3 is a protein-targeting factor, analogous to the signal recognition particle (SRP), and, similar to SRP components (Hann and Walter, 1991), is not essential for viability in yeast; however, the cells are sensitive to a variety of stresses such as heat and metals (Zúñiga et al., 1999).

Get3 contains a nucleotide hydrolase domain (NHD) that resembles the G-type hydrolases characterized by Ras (for review see (Sprang, 1997)). These proteins all have the completely conserved 'P-loop' that recognize the α - and β -phosphate in both NDP and NTP states. Other features of G-type hydrolases are Switch I (A') and Switch II (Walker B) loops that undergo dramatic rearrangements coupling structural changes to the presence of the γ -phosphate. In these proteins, catalysis is stimulated by a positively charged residue that stabilizes negative charge on the phosphates and a residue that positions a catalytic water for nucleophilic attack.

Get3, like ArsA and the nitrogenase iron protein (NifH), belongs to a special class of ATPases that contain a 'deviant' Walker A motif which is a P-loop with an additional invariant lysine (G**K**GGVGKT in Get3) (Koonin, 1993). This is a rare motif, found in only two other yeast proteins (including a putative Fe-protein homologue (Netz et al., 2007)). A basic model for the deviant P-loop ATP hydrolysis cycle can be inferred by the structure of a NifH dimer bound to ADP·AlF₄⁻ and its partner the MoFe protein (Schindelin et al., 1997). The ADP and apo forms of NifH are in an open conformation that is inactive for ATP hydrolysis (Schlessman et al., 1998). Binding of the MoFe protein, along with ATP, causes a large rotational and translational shift of the two NifH monomers that brings the deviant P-loop lysine from the opposing monomer into a position to stabilize the build up of negative charge on the phosphates. This is analogous to the mechanism in Ras where an Arg-finger from a GAP

stimulates hydrolysis of ATP leading to Ras inactivation (Milburn et al., 1990; Scheffzek et al., 1997). This interface shift demonstrates how ATP can modulate dramatic structural changes. Critical to all of this, the rearrangements are stabilized by binding of the MoFe protein (Schindelin et al., 1997). In the case of ArsA, without its partner ArsB bound, no states are found in which both NHD bind the same nucleotide and it is reasonable to speculate that in a true ATP state a dramatic conformational change must occur as well (Zhou et al., 2001; Lutkenhaus and Sundaramoorthy, 2003).

There are no mechanistic studies detailing how Get3 performs its important targeting function and a molecular level understanding requires structural information. Here we present three crystal structures of Get3/TRC40, a monomeric apo form from *Saccharomyces cerevisiae* (*Sc*Get3) and dimeric apo and hexameric ADP-bound forms from the thermophilic opportunistic human pathogen *Aspergillus fumigatus* (*Af*Get3 and *Af*Get3-ADP). Based on the structures, we probed functional interfaces and essential residues by phenotypic rescue. Our results allow us to define a model of how Get3 couples ATP hydrolysis to the binding and release of TA proteins. More broadly, this work supports a mechanism for a special class of ATPases.

Results

Crystallization of Get3

We purified *Sa*Get3 and *Aj*Get3 from constructs expressed in *E. coli* using Ni-affinity and size exclusion chromatography under reducing conditions. The majority of the protein eluted as a dimer from both constructs and this was used in crystallization trials. The *Aj*Get3-ADP crystals diffracted to 3.2Å resolution in the space group $P2_12_12_1$ with a hexamer in the asymmetric unit. A seleno-methionine data set was collected to 4.5Å resolution and phases were solved by multiwavelength anomalous dispersion and extended to 3.2Å resolution using the 6-fold noncrystallographic symmetry. The final refined structure contained 292 of 348 residues with a Free-R factor of 25.1% (Fig. 1C and Fig. S1). The two Apo forms of Get3 were solved using molecular replacement of a truncated *Aj*Get3-ADP monomer. The *Sa*Get3-apo crystal form diffracted to 3.7Å resolution in the space group $H32$ and contained a monomer in the asymmetric unit. The final model contained 260 of 369 residues and refined to a Free-R factor of 33.5% (Fig. S2 A and B). The *Aj*Get3-apo crystal form diffracted to 7.5Å resolution in the space group $P4_232$ and contained a dimer in the asymmetric unit that we did not refine due to the low resolution (Fig. S2D). Crystallographic statistics are provided in Table S1.

Description of a Get3 Monomer

The structure of a monomer of Get3 is a mixed alpha-beta fold containing a “P-loop” type NHD with two α -helical loops that extend outward from the structure, here designated substrate binding loop 1 (SB1) and 2 (SB2) (Fig. 1 A and B). The Get3 NHD fold falls into a more specific structural class (defined by SCOP (Andreeva et al., 2008)) that includes ArsA (Zhou et al., 2000), GTPase domains of the signal recognition particle (SRP) (Freyman et al.,

1997), and SRP receptor (Montoya et al., 1997) along with NifH (Georgiadis et al., 1992). Get3 is the only eukaryotic example in this class that utilizes ATP.

The Three Crystal Forms of Get3

The *Af*-ADP crystal form contains a hexamer in the asymmetric unit (Fig. 1C) with 3-fold symmetry in which the monomer can be assembled into two potential dimers formed by either SB1/2 (arm dimer) or by the interface between the NHD (Fig. 1 B and D). Although the arm dimer contains a more extensive interface, 1758Å² versus 1263Å² calculated by PISA (Krissinel and Henrick, 2007), we believe that the NHD dimer is the most relevant to TA protein binding and contains two disulfides formed across the interface by a conserved pair of cysteines (Fig. 1B).

The *Sc*-apo crystal form contains a monomer in the asymmetric unit (Fig. 1E and Fig. S2A) and there is no apparent NHD dimer interface despite the protein purifying as a dimer. The crystals are stabilized by the coordination of a zinc by Cys-285/Cys-288 and the His-tag which also coordinates a second metal at the crystallographic three fold in a square-planar geometry (Fig. S2A). The *Af*-Get3-apo contains two copies in the asymmetric unit in an orientation similar to the NHD dimer, although rotated so that the SB1 and SB2 regions of the two monomers are slightly closer (Fig. S2E). Both Apo crystal forms contain symmetry related interfaces similar to the arm dimer (Figs. S2C and S3F).

Monomers in each of the three crystal forms show slight variations in SB1/2 demonstrating the general mobility of these regions (Fig. 1E and Fig S2F). These loops in the hexamer are stabilized by a series of hydrophobic interactions in a highly flexible region that perhaps explains the difficulty in obtaining high-resolution crystals. SB1 and SB2 contain disordered regions in all of the crystal forms. The missing residues of SB1 have been modeled into the

A/Get3-ADP hexamer to demonstrate the amount of disordered protein that could not be built (Fig. 1C).

Nucleotide Binding

The Get3 nucleotide-binding pocket contains all of the features generally found in G-type hydrolases. The completely conserved Asn in S7 (*Sc*/*A*/272) forms hydrogen-bonds that specifically select for adenine. Additional interactions with the A-loop complete adenosine recognition (Fig. 2A). The P-loop, as is typical, makes extensive contacts to the α - and β -phosphate; however, the second lysine, completely conserved in P-loops, is in an orientation that points away from the β -phosphate. This is caused by an interaction in the arm dimer that leads to an Arg from SB2 (*A*/200) moving into the active site occupying a similar position near where one would expect Mg^{2+} to be bound (Fig. 2 A). It is clear that *A*/R200 displaces the Mg^{2+} and generally disrupts the interactions of Switch I and II. Based on the resolution, we cannot be certain that there is no Mg^{2+} ; however, if present it would be in a unique position. *A*/R200 forms a salt bridge to the β -phosphate; but it is not conserved making the extent of these interactions surprising.

Comparison to ArsA and NifH

Despite distinct functions, Get3 shares a similar topology to ArsA with an RMSD of 1.9Å in their NHD (Fig. 2B) (PDBID 1f48) (Zhou et al., 2000). In contrast to Get3, ArsA SB1/2 bend in across the NHD dimer interface forming a coordination site for heavy metals (Fig. 2A); however, these coordinating residues are not conserved in Get3. It is thought that motions of these loops are coupled to ATP hydrolysis regulating metal release via the Switch II motif (Zhou et al., 2001). The dimer interface is very similar to Get3 except that the interface is rotated moving the P-loop from 9.1Å (G17/G336) separation in ArsA to 14.1Å (*A*/G35) in

Get3. An early homology model of Get3, based on ArsA, predicted the occurrence of the disulfide bridges between the subunits at the dimer interface. Based on the model they found that mutation of the two cysteines in Get3 was unable to rescue a metal sensitivity phenotype in a Get3 knockout (Metz et al., 2006). ArsA is a pseudodimer with a disordered linker peptide between the two subunits that may be required to stabilize the dimer interface.

The best understood deviant P-loop protein is NifH as its structure has been solved in Apo, ADP, and ADP·AlF₄⁻ forms. The structure closest to *A. vinelandii* Get3-ADP is the NifH-Apo form (Schlessman et al., 1998) and the NHD domains have an RMSD of 2.78Å. As noted, ATP stimulates a large conformational shift that moves the deviant P-loop (*A. vinelandii* NifH G11) from 10.1Å to 4.0Å apart. To move the Get3 dimer into a similar orientation would require an extensive conformational change across the dimer interface.

Tail-Anchor Binding Pocket

In a search for the TA protein binding pocket, the positions of SB1 and SB2 are clearly provocative. We analyzed the NHD dimer by displaying conserved and hydrophobic residues on a molecular accessibility surface (Fig. 3). The interface, formed by the NHD dimers, is highly conserved, as is expected for a common fold (Fig. 3A). The other concentration of conserved residues is found at the base and groove formed by SB1/SB2 (Fig. 3C). The overall alignment of SB1/SB2 is difficult; however, there is general conservation of hydrophobic residues and glycines (Fig. S1). Additionally, SB1 contains a disordered stretch that could also provide surface area in this region. The only hydrophobic patch on the surface of the dimer is that formed between SB1 and SB2 (Fig. 3 B and D). This putative TA-binding region would be analogous to the location of the metal binding sites of ArsA or the Fe-S cluster in NifH (Fig. 2

B and C) and one presumes that changes in the ATP binding pocket would be transmitted to this region.

Phenotypic Rescue

To probe the functional parts of Get3 we chose a series of mutants based on surface conservation or putative function and tested for their ability to rescue known knockout phenotypes (Metz et al., 2006). The knockout ($\Delta get3$) showed no obvious impairment on synthetic complete media at 30C but was unable to fully rescue on media containing Cu^{2+} or hydroxyurea or growth at elevated temperature. Replacing the Get3 gene on a plasmid (*GET3*) with the wild-type promoter rescued the $\Delta get3$ growth to near wild-type levels. We also inserted the *Af*Get3 gene on the same plasmid and this also rescued the yeast knockout demonstrating that functional aspects of the protein are conserved across species (Fig. 4A and Fig. S3). In all, we generated 69 *Sc* mutants and two *Af* mutants and scored their general loss-of-function (LOF) phenotype as strong, moderate, weak or none (Fig. 4A and B, data for all mutants in Fig. S3 and Table S2).

In agreement with previous results, a mutation in either the P-loop (*Sc*G30R or *Af*G38R) or the pair of Cys that form the disulfide bridges (C285T/C288T) had strong LOF phenotypes (Fig. 4A) (Schuldiner et al., 2005; Metz et al., 2006). The G30R mutation is thought to disrupt ATP binding. The effect of the Cys mutants is less clear. Presumably, this interface is somewhat unstable and requires the disulfides to stabilize the dimer, similar to the linked dimer of ArsA. As the cytoplasm is a reducing environment, it would be curious that the disulfides could form in vivo; however, we included reducing agent in all of our buffers and the disulfides formed in that context. Another possibility is that these residues coordinate a metal

or are regulated by a redox pathway (Metz et al., 2006) as the reduced form in the *Sc*Get3-apo crystals is a monomer and coordinates a zinc (Fig. S2 A and B).

The largest cluster of LOF mutants occurs at the NHD dimer interface found mostly on H8 and H9. The interface is a mixture of hydrophobic and charged groups that would be intimately involved in a re-arrangement of the dimer (Fig. S4B, Figs. S1 and S4A). Only a few of the conserved surface mutations that did not make contacts in this crystal form conferred LOF phenotypes (R75A, D265A, and Y338A) and probably do not affect the conformational changes in substrate binding (Fig. S4B). It is possible that these surface residues play a role in recognition of other proteins in the Get pathway.

Changes in switch helices are normally coupled to functional changes. Although there was little conformational change in ArsA nucleotide structures, it was postulated that binding of ATP would cause a conformational change in Switch II that would be transmitted to a His involved in metal coordination (Zhou et al., 2001). This residue (*Sc*H172) is the only coordinating residue from ArsA that is found in Get3 and in our structure this is in a position to interact with a network of salt bridges that appear to stabilize the base of SB1/2 (Fig. S4C). Mutations of these residues had LOF phenotypes; however, they were not strong and it may be possible that coupling of Switch II changes is not essential for TA binding.

Get3 binds a variety of TA protein substrates via hydrophobic interactions (Stefanovic and Hegde, 2007) and it is difficult to decide what mutations might interfere with binding. Based on the predicted pocket, we generated extensive mutations in SB1 and SB2. As expected, mutations in the predicted TA protein binding pocket (I136S, D137A, L140S, S141A, M143S, and L219S) had LOF; however, the majority of the residues had no phenotype including those

disordered in our structure (Fig. 4B and Fig. S4D). This hydrophobic interaction may require multiple mutations in the binding pocket to see significant disruptions.

The lack of conservation of the SB2 residue *A/R200* made a comparable mutation in *Sc* impossible. Due to its location in the nucleotide-binding pocket in the *A/Get3*-ADP structure we decided to see if its mutation would have an effect on rescue by *A/Get3*. An *A/R200A* mutation was a clear LOF phenotype (Figs. 2A and 4A). This is in contrast to a number of other mutations, including some disordered residues, in this region of SB2 that showed no phenotypes. It is difficult to envision an effect of the *A/R200A* mutation in the absence of the hexamer.

A mutation of the conserved deviant P-loop lysine is expected to completely abolish function and should be a strong phenotype. In the *A/Get3*-ADP structure this residue makes no contacts (Fig. 4 and Fig. S4A), still, a mutation of this residue to Ala (*ScK26A*) was the strongest phenotype of all. In the Ras/ RasGAP case any mutation of the Arg-finger leads to a total LOF even for the seemingly benign mutation to Lys (Ahmadian et al., 1997). We did the same type of mutation, *ScK26R*, and found that this mutant is a strong LOF phenotype, although not as strong as *ScK26A*.

Figures

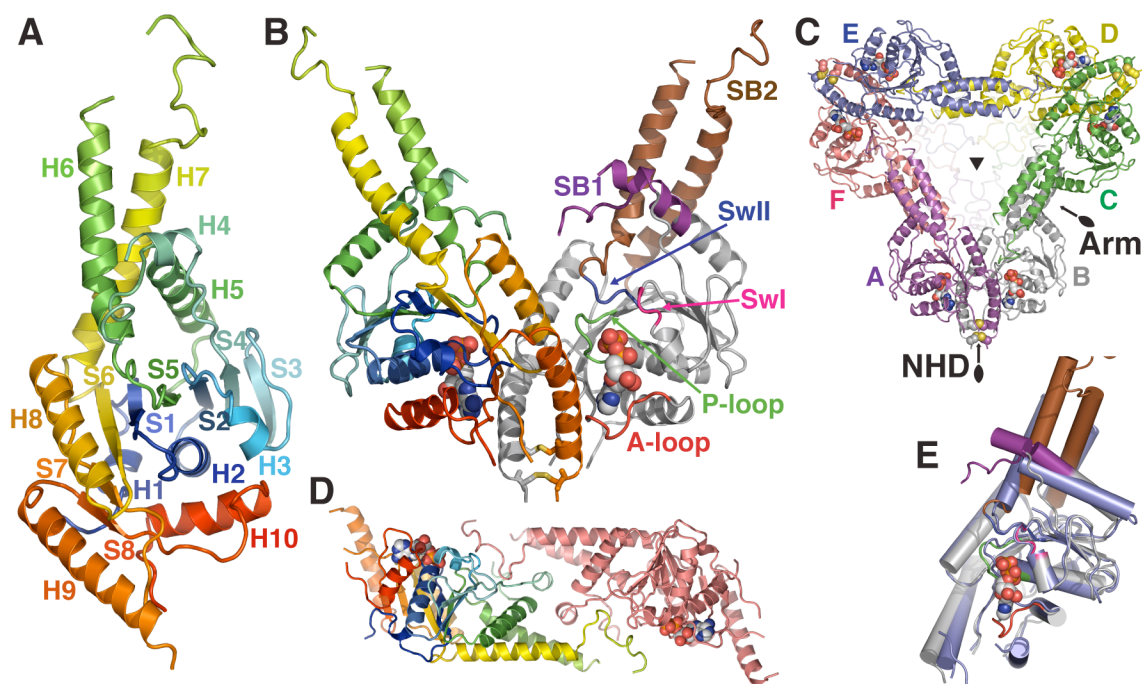


Figure 1. Structures of Get3. (A) An *A/Get3*-ADP monomer with secondary structure elements numbered as in Fig. S1 (B) The NHD dimer of the *A/Get3*-ADP hexamer. One monomer is color ramped from N- (blue) to C- (red) and the other is colored relative to motifs described in the text: P-loop (green), Switch I (magenta), Switch II (blue), A-loop (red), SB1 (purple), and SB2 (brown). (C) The *A/Get3*-ADP hexamer of the asymmetric unit colored by monomers. Dimer interfaces and subunits are labeled. The 3-fold is indicated by a triangle and 2-folds are indicated by an oval. The modeled residues from SB1 of each monomer are transparent. Nucleotides and bridging disulfides are shown as spheres. (D) The arm dimer of the *A/Get3*-ADP hexamer. One monomer is color ramped as in A and the other is colored salmon. (E) The *ScGet3* model colored purple and overlaid with an *A/Get3*-ADP monomer colored and oriented as in B.

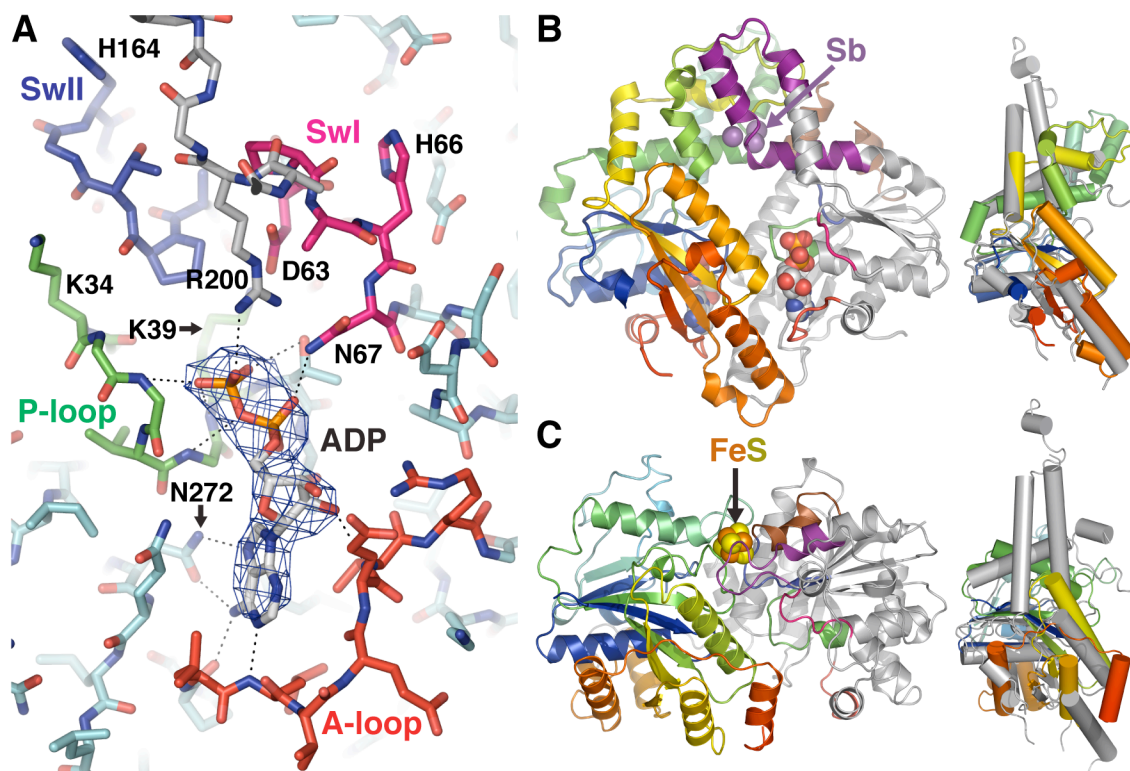


Figure 2. The nucleotide-binding pocket and comparison of Get3 to other hydrolases. (A) The nucleotide-binding pocket of *A/Get3*-ADP with residues shown as sticks. Density is a $2F_o - F_c$ omit-map contoured at 1.5σ . (B) A ribbons diagram of the ADP form of *EcArsA* (1f48) with ADP, Mg^{2+} (green) and coordinated Sb (purple) as spheres. (C) A ribbon diagram of the apo form of NifH (2nip) with the Fe/S cluster (orange/yellow) as spheres. To the right in B and C are overlays of monomers the *A/Get3*-ADP monomer (gray) on the respective left subunit. Important residues and motifs are labeled. All residues in nucleotide binding motifs are colored as in 1B.

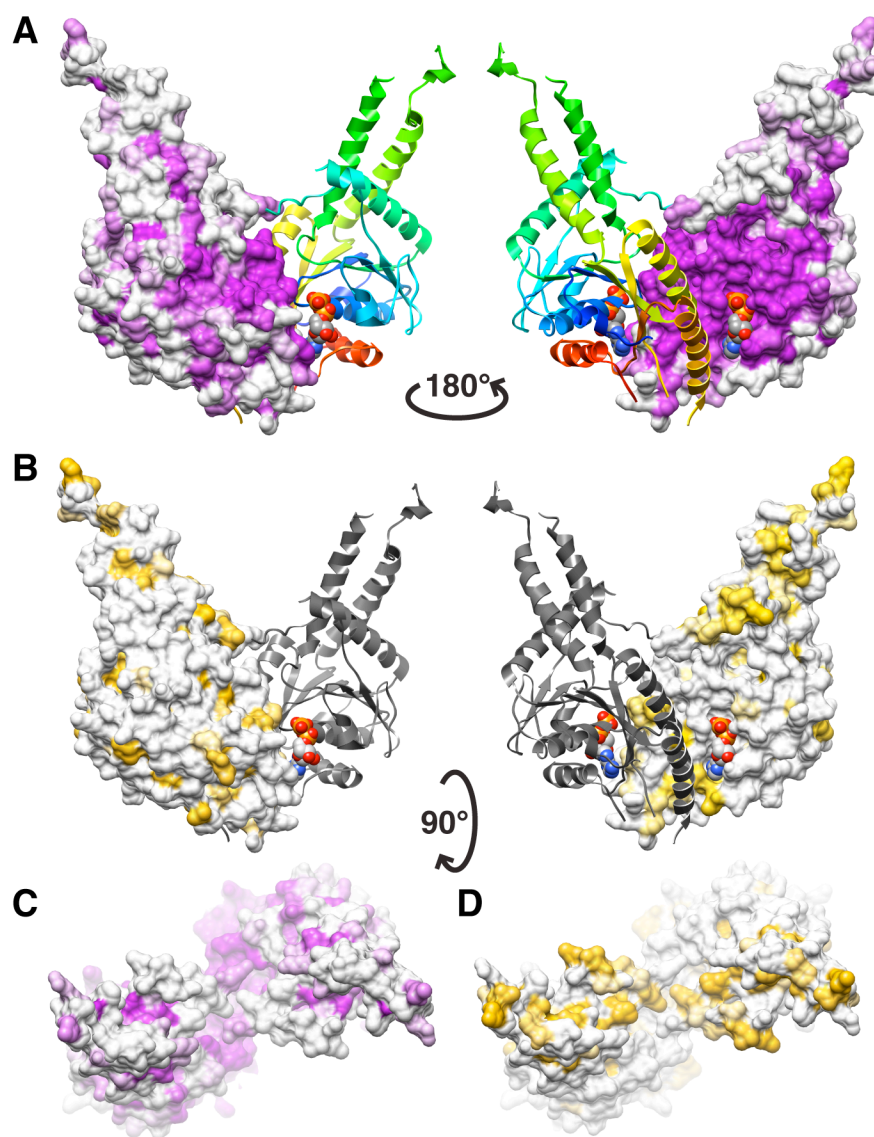


Figure 3. Conserved and hydrophobic surfaces. (A) Front and back views showing one monomer as a ribbon colored as in 1A and the other as an accessible surface showing conservation colored as a gradient from 100% (purple) to 50% conserved (gray). Conservation is based on the Get3 alignment from Fig S1 (B) Similar to A showing hydrophobicity based on the Kyte and Doolittle scale with most hydrophobic in dark yellow. (C) Conserved surface viewed from the top. (D) Hydrophobic surface viewed from the top.

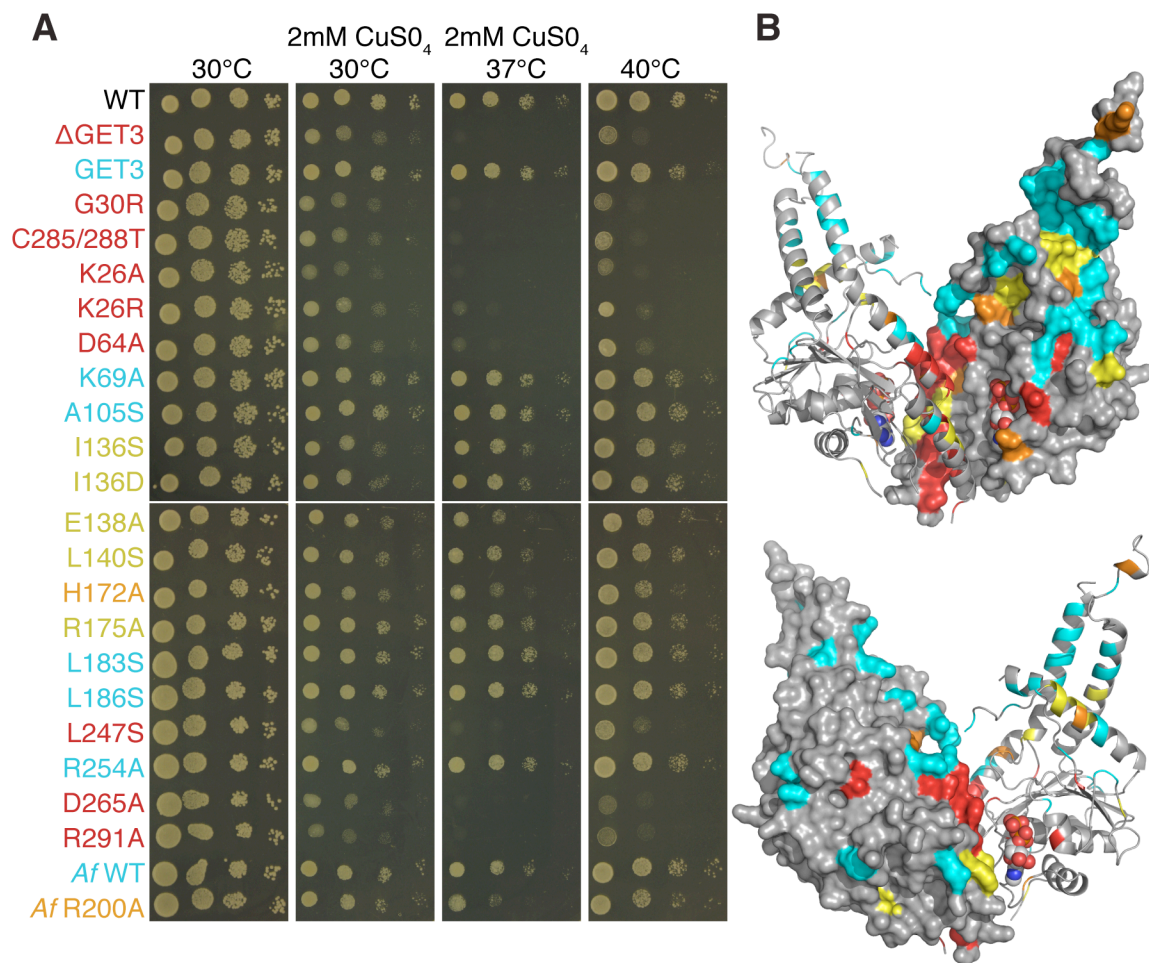


Figure 4. Phenotypic rescue of various Get3 mutants. (A) Spot plate assays of various yeast mutants on a plasmid with a wild-type promoter screened on SC-Ura plates at 30C and 40C and supplemented with 2 mM CuSO₄ at both 30C and 37C. ΔGET3 is the knockout transformed with a plasmid containing only the promoter. Mutants based on yeast numbering are indicated colored relative to their phenotype: strong (red), moderate (orange), weak (yellow), and none (cyan). (B) Two views with one monomer as accessible surface and the other in ribbon colored according to phenotype as in A.

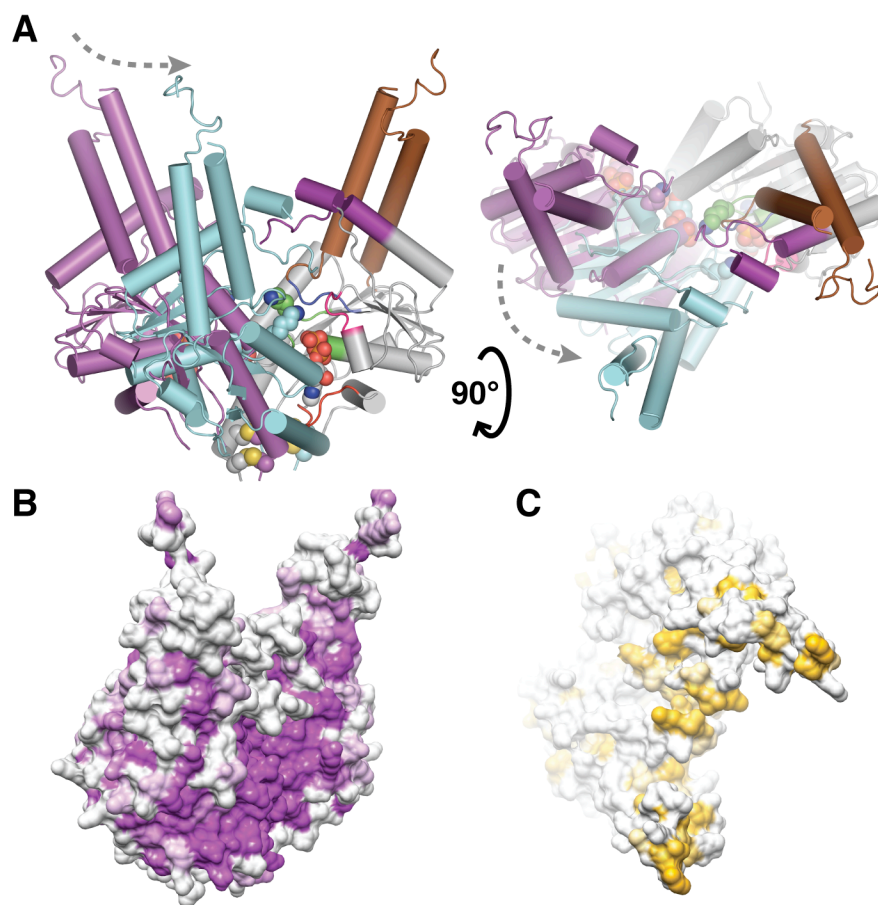


Figure 5. NifH-like model. (A) An overlay of *A/Get3*-ADP in the open dimer and the closed NifH (1m34)-like model. The *A/Get3*-ADP dimer, similar to Fig. 1B, is colored with the right monomer by feature and the left in purple. The modeled rotated monomer is in light blue. *A/K34* and the bridged cysteines shown as spheres. Arrows indicate direction of motion. (B) Conserved surface of the NifH-like model oriented as left A. (C) Hydrophobic surface of the NifH-like model orientated as right A.

Discussion

Get3 must couple ATP hydrolysis to TA protein binding and release. To propose a mechanism for Get3 binding of TA proteins we can model the transition to a closed-bound state based on the NifH structures. We believe our structures represent various open, non-substrate binding conformations where the hydrophobic SB1/2 are highly flexible and open for interaction with proteins, possibly in a metastable hexameric state. The binding of ATP couples to a rearrangement of the Switch loops that would be transmitted to SB1/2. Binding would also involve a rotation and translation at the dimer interface that moved the bridging *S*K26 (*A*/34) into a position to counter the additional charge of the γ -phosphate (modeled in Fig. 5A). The *A*/Get3-apo dimer shows how some of this motion could occur as it rotates inward, relative to the *A*/Get3-ADP dimer, demonstrating flexibility at this interface (Fig. S2E). There are clashes in this simple NifH-like model and we believe that additional conformational changes must occur.

This ATP bound complex would bury a considerable amount of the conserved residues at the dimer interface (Fig. 5B) and would bring SB1/2 from opposing dimers into a closer orientation, creating a large hydrophobic groove at the top of the interface (Fig. 5C). This structure would be incompatible with our hexamer but would provide a favorable binding surface for a TA protein. The TM helix would dock in the groove formed at the base of SB2 and the hydrophobic flexible loop of SB1 could then wrap around it, similar to SRP signal sequence finger-loop binding (Schaffitzel et al., 2006). The only component missing in such a model is the residue that would activate the catalytic water. It is possible that additional partner binding at the membrane would either donate this group or lead to additional conformational changes in Get3 that would stimulate ATP hydrolysis once the substrate has been delivered.

The oligomeric state of Get3 based on this work leads to open questions about function. By homology to NifH and ArsA, we have described a model in which the NHD dimer interactions are the most relevant to TA protein binding; however, we find it difficult to ignore the arm dimer interface. In all of our crystal structures, SB1/2 interactions bury a significant amount of hydrophobic surface implying that they have a high affinity for binding protein (Fig. 1D and Fig. S2 C and F). In an open form these surfaces should be very unstable and it is hard to imagine that they could exist free in the cytoplasm. The hexamer seen in our crystal structure could be a stable resting form of the protein that needs additional factors, such as the Get4/Get5 proteins (Jonikas et al., 2009), to transition to the open dimer state. Another possibility is that the hexamer operates as an ADP-exchange factor (like a GEF for Ras) stabilizing the apo form for ATP binding by displacing the Mg^{2+} and releasing ADP. In *Aj*/Get3 the R200 salt bridges to the ADP β -phosphate, which would seem to stabilize the ADP form; however, the concentration of ADP in our crystal conditions is very high and the binding could be an artifact of that. A third, less likely, possibility would be that the hexamer is the active form of the complex and that TA proteins are stabilized in the flexible hydrophobic center reminiscent of some AAA ATPases (Ogura and Wilkinson, 2001). Evidence that supports a role for the hexamer is the importance of *Aj*/R200, a purified human Get complex sediments at a compatible size (Stefanovic and Hegde, 2007), the functional form of ArsA is a multimer (Ching et al., 1991), and a trimeric form of ArsA has been visualized by EM and chromatography (Wang et al., 2000).

Proper synthesis and targeting of TA proteins by the Get pathway have broad implications in biology, as they are essential in many cellular homeostasis and transport processes. Our structural and functional studies are a mechanistic look at the recently identified pathway component Get3. These experiments allow us to define a model that predicts the

conformational changes in Get3 that are involved in TA protein and nucleotide binding (Fig. S5). They also suggest an oligomeric form that may play a key role. Many aspects of TA protein targeting, such as the specifics of substrate binding, interactions of partners and the kinetic steps of recognition and release, remain to be determined.

Materials and Methods

Cloning, Expression, and Purification

The *A. fumigatus* Get3 coding sequence was synthesized by PCR using primers designed with DNABWorks (Hoover and Lubkowski, 2002), and the *S. cerevisiae* GET3 gene was amplified by PCR from genomic DNA, both with NcoI and XhoI restriction sites added. Amplified DNA fragments were NcoI/XhoI digested and ligated into pET33b (Novagen) to create C-terminally 6×His tagged constructs.

A. fumigatus and *S. cerevisiae* Get3 proteins were recombinantly expressed in *E. coli* BL21-Gold(DE3) cells grown in 2× YT medium for 3 h. at 37 °C after induction with 0.3 mM isopropyl-β-D-thiogalactopyranoside. Get3 was purified by Ni-NTA affinity chromatography and gel filtration on a Superdex 200 16/60 column (GE Healthcare). Fractions from gel filtration were concentrated to 10–15 mg/mL of protein and dialyzed in a buffer of 5 mM Tris, pH 7.5, and 6 mM beta-mercaptoethanol (βME) for crystallization.

Crystallization and Structure Determination

All crystals were grown by the sitting-drop vapor diffusion method with a 1:1 ratio of protein to precipitant solutions at 23°C. *Sa*Get3-apo crystals were obtained in 0.1 M HEPES pH 8.0, 1.6 M ammonium sulfate and 6 mM βME, *Aj*Get3-apo crystals in 0.1 mM bis-Tris, pH 7.0, 0.2 M NaCl, 1.5 M ammonium sulfate and 6 mM βME, and *Aj*Get3-ADP co-crystals in 0.2 M potassium citrate, 16% (wt/ vol) polyethylene glycol 3350, and 6 mM βME, with the protein solution supplemented with 2 mM ADP and 1 mM MgCl₂. *Sa*Get3-Apo were cryoprotected by transfer to 3.4 M sodium malonate (pH 7), and *Aj*Get3-Apo crystals and *Aj*Get3-ADP co-crystals were serially transferred to artificial mother liquor supplemented with 20% (wt/vol) sucrose and 20% (wt/vol) xylitol and with 20% ethylene glycol, respectively. All crystals were

flash-frozen to 100 K by direct transfer into liquid nitrogen. Selenomethionine (SeMet) derivatives of *Aj*Get3 were handled in the same manner as the native protein.

All diffraction data were obtained on Stanford Synchrotron Radiation Lightsource beam line 12–2 at the SLAC National Accelerator Laboratory, at 100 K. (Table S1). Diffraction data were integrated with MOSFLM and scaled with SCALA (Leslie, 1992; Evans, 2006). Multiple-wavelength anomalous dispersion (MAD) data from a SeMet derivative of *Aj*Get3-ADP allowed the assignment of forty-nine selenium sites using SHELXD, and experimental phases were calculated by SOLVE with an overall figure of merit of 0.63 (Terwilliger and Berendzen, 1999; Sheldrick, 2008). BUCCANEER performed density modification and built an initial model, and the complete model was manually built in COOT (CCP4, 1994; Emsley and Cowtan, 2004).

Refinement against the 3.2Å resolution native *Aj*Get3-ADP data used strict 6-fold NCS symmetry and consisted of cycles of simulated annealing and group B-factor refinement in CNS followed by manual rebuilding (Adams et al., 2002; Painter and Merritt, 2006a). Unambiguous density was observed for residues 12–106, 125–189, 195–277, and 282–338 in all monomers. The overall topology is shown in Fig. S2A. Density was observed for residues 190–194 in the SB2 loops but could not be confidently modeled. Residues 195–200 were modeled individually into each monomer. TLS groups were determined using the TLSMD web server and NCS restraints were relaxed to allow variation between subunits for final refinement in PHENIX to yield an R-factor of 21.2% and an R-free of 25.13% (Adams et al., 2002; Painter and Merritt, 2006a).

Molecular replacement of the *Sc*Get3-apo dataset was performed with PHASER (McCoy, 2007). The search model was prepared from the *Aj*Get3 ADP structure by removing SB1 loop

residues 98 to 129, SB2 residues 177 to 214, and residues 282–288 of H9. A single copy was located in the asymmetric unit and the initial weighted $2 F_o - F_c$ maps showed density that allowed rebuilding of H9, an extension of the C terminus by 11 residues to form a helix that packs against S8, and shifts in H5, H6, and H7 (Fig. 1E and Fig. S1A).

An anomalous difference map contained a strong peak between C285 and C288 and a second peak at the proximal 3-fold axis special position. X-ray fluorescence spectroscopy indicated that the crystal contained zinc. Density between the two peaks could be modeled as the C-terminal hexahistidine tag used in purification as an extended strand, resulting in square planer coordination of a metal between the two cysteines and two histidines, and octahedral coordination at the special position by symmetry-related histidines.

After a manual rebuild, the helices and the β -sheet were refined as rigid bodies and one isotropic B-factor was refined per residue. Hydrogen bond and ϕ/ψ angle restraints allowed for torsion angle simulated annealing in CNS and PHENIX while preserving secondary structure geometry. The refined model consisted of residues 8–89, 137–190, 217–277, 284–316, and 320–356 with an R-factor of 28.1% and an R-free of 33.5%.

The truncated *A/Get3*-ADP search model was also used for molecular replacement of the *A/Get3*-apo data. Density was observed in the weighted $2 F_o - F_c$ maps corresponding to the SB2 region (Fig. S2D). Two copies were located in the asymmetric unit in a relative orientation similar to the NHD dimer although rotated so that the SB1 and SB2 regions of the two monomers are slightly closer (Fig. S2E). Symmetry related copies form dimers similar to the arm dimer (Fig. S2F). No further refinement of the apo structure was performed.

Figs. 1, 2, 4, and 5A, Figs. S2, and S4 were prepared using PyMOL (Schrödinger, LLC). Figs. 3 and 5 B and C were prepared using UCSF Chimera (Sanner et al., 1996; Pettersen et al., 2004).

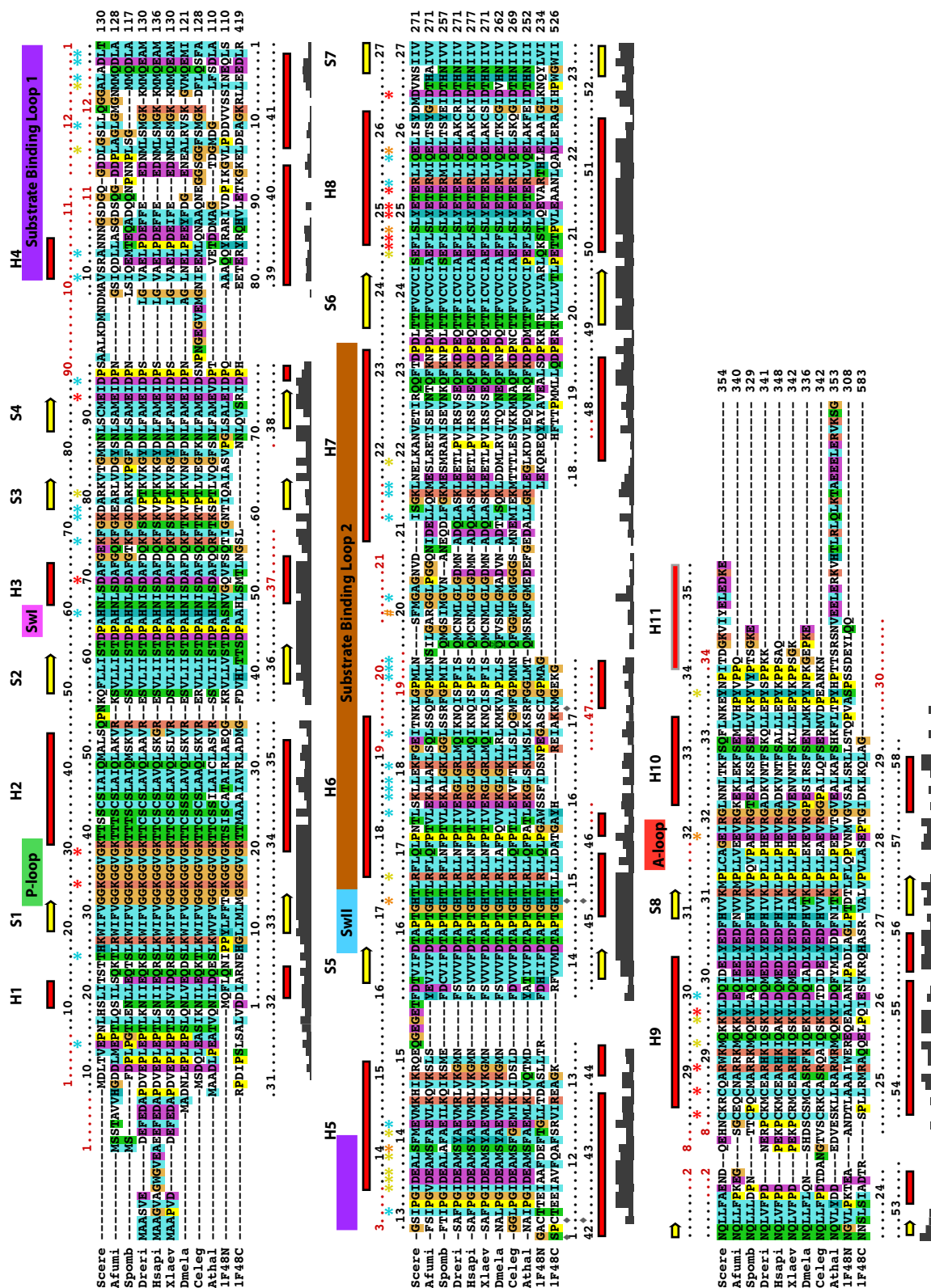
Growth Assays

The promoter region of *GET3* was amplified from genomic DNA by PCR with *Xba*I and *Nco*I restriction sites and ligated 5' to *A. fumigatus* and *S. cerevisiae* *GET3* in the pET33b constructs. The promoter region, promoter region with *S. cerevisiae* *GET3* and promoter region with *A. fumigatus* *GET3* were each amplified by PCR with *Xba*I and *Nco*I restriction sites added and cloned into YEp352 vector (ATCC). *GET3* mutants were generated by site-directed mutagenesis. YEp352 constructs were transformed into BY4741 and BY4741 *YDL100c::kanMX4* cells (ATCC) for use in growth assays.

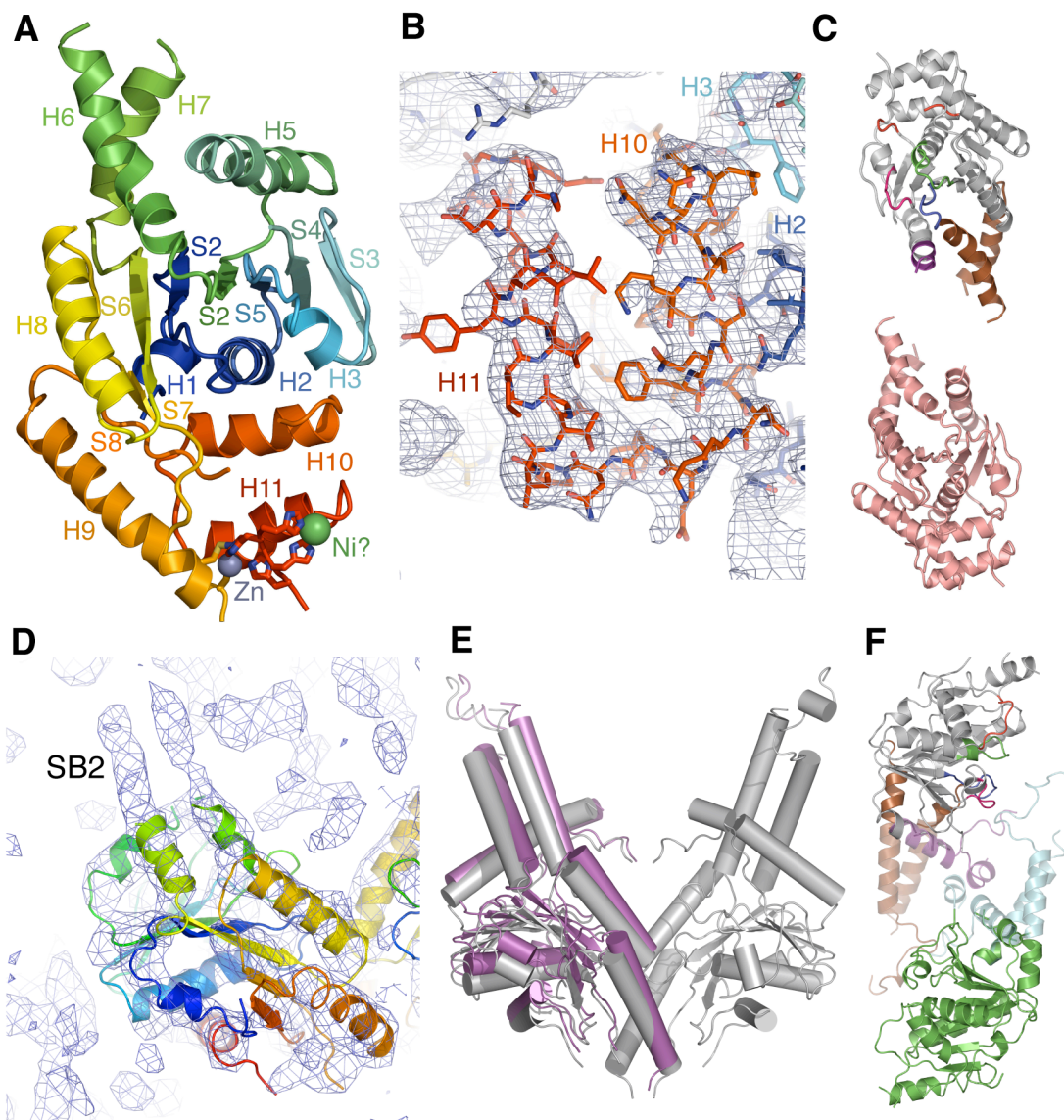
Growth defects of *GET3* knockouts complemented with *GET3* mutants in YEp352 on drop plates were scored were given one of three scores: weak, moderate and strong. Growth defects weaker than the knockout were categorized as weak if they were closer to the WT and moderate if they were closer to the knockout. Growth defects similar to or greater than the knockout were classified as strong. We determined the consensus score by taking the strongest growth defect among the different growth conditions and averaging it between the duplicates from separate experiments.

ACKNOWLEDGMENTS

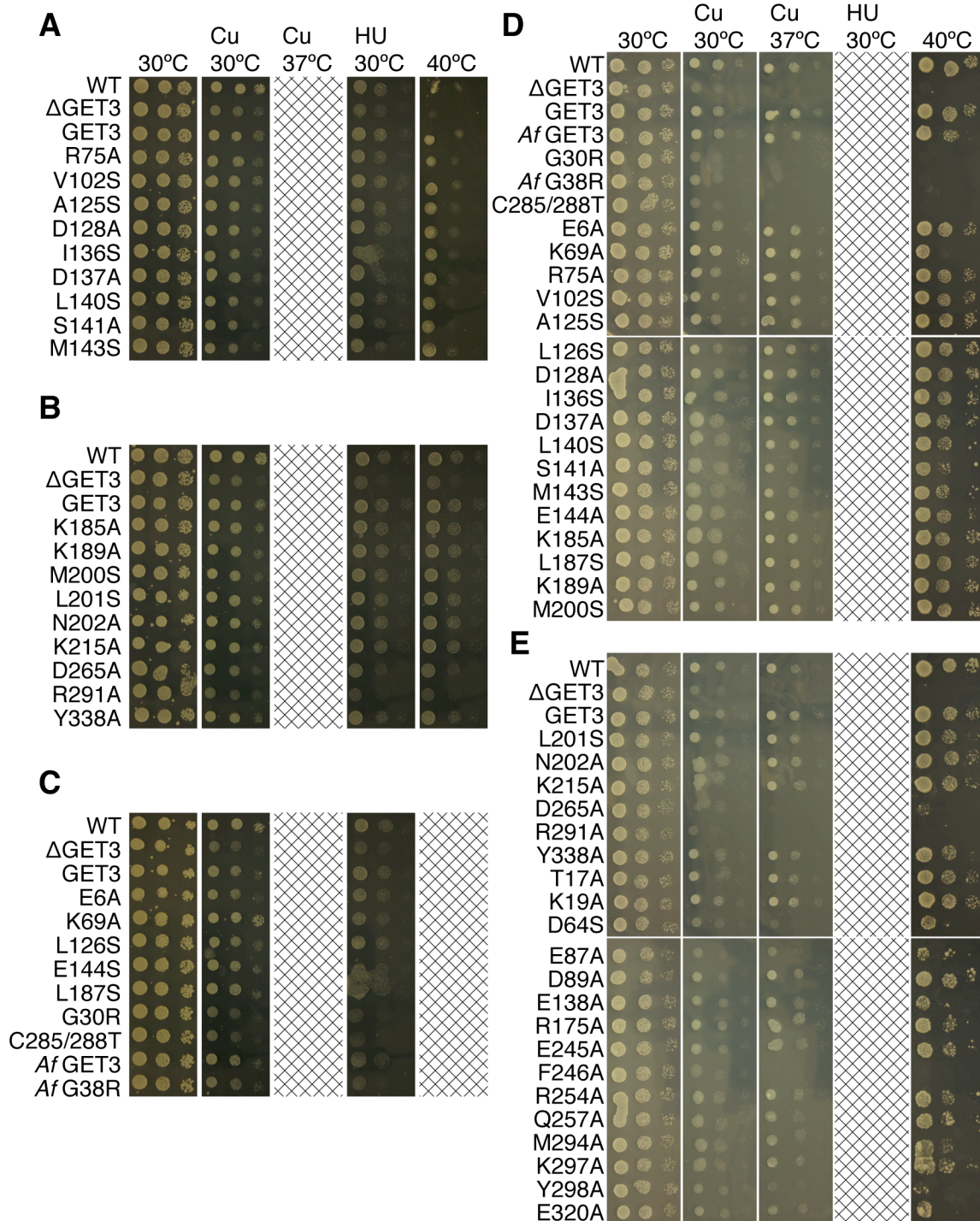
We thank S. Shan and A. Palazzo for discussion and critical comments on the manuscript and Gordon and Betty Moore for support of the Molecular Observatory at Caltech. All data collection was performed at beamline 12-2 at Stanford Synchrotron Radiation Lightsource and we thank G. Card and M. Soltis for beamline assistance. Operations at SSRL are supported by the U.S. Department of Energy and National Institutes of Health. W.M.C. is supported by the Searle Scholar program and a Burroughs-Wellcome Fund Career Award for the Biological Sciences.

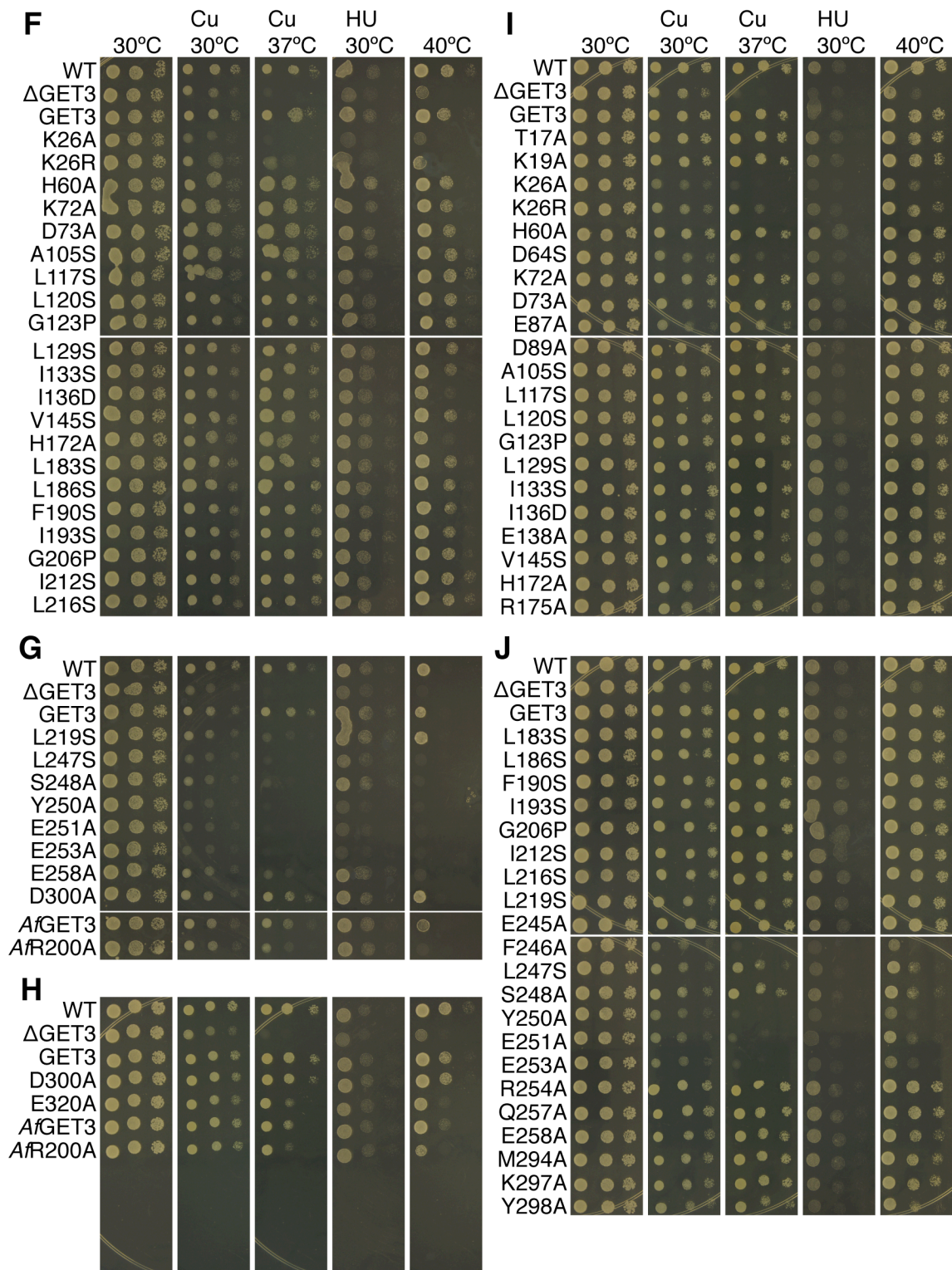


Supplemental Figure 1. Sequence alignment of Get3 homologues. Sequences were aligned using the program ClustalX (Larkin et al., 2007). Residue coloring is based on the program output (colored based on amino acid type). The species in order are Scere (*S. cerevisiae*), Afumi (*A. fumigatus*), Spomb (*Schizosaccharomyces pombe*), Drerio (*Danio rerio*), Xlaev (*Xenopus laevis*), Dmela (*Drosophila melanogaster*), Celeg (*Caenorhabditis elegans*), Athal (*Arabidopsis thaliana*), and 1F48N and C (the N and C-terminal sequence of *E. coli* Arsa). Numbering, from top to bottom, is based on Sc, Af, and EcArsa with disordered residues in ScGet3, AfGet3-ADP and PDBID 1f48 colored in red. Secondary structure for AfGet3-ADP is shown on top, along with numbering as in Fig. S2 A, and the N terminus of 1f48 is shown on the bottom, α -helices as red rectangles and β -sheets as yellow arrows. Structural elements discussed in the text are in boxes above the alignment colored as in Fig. 1B. Below the alignment, gray bars show degree of conservation at a given position based on Get3 sequences. Mutations described in the text are indicated by asterisks for Sc (*) and pound for Af (#) colored strong (red), moderate (orange) and weak (yellow) for LOF phenotypes and cyan for mutants that did not display a phenotype in our assays. Gray diamonds (filled diamonds) represent residues in Arsa that coordinate metal binding.

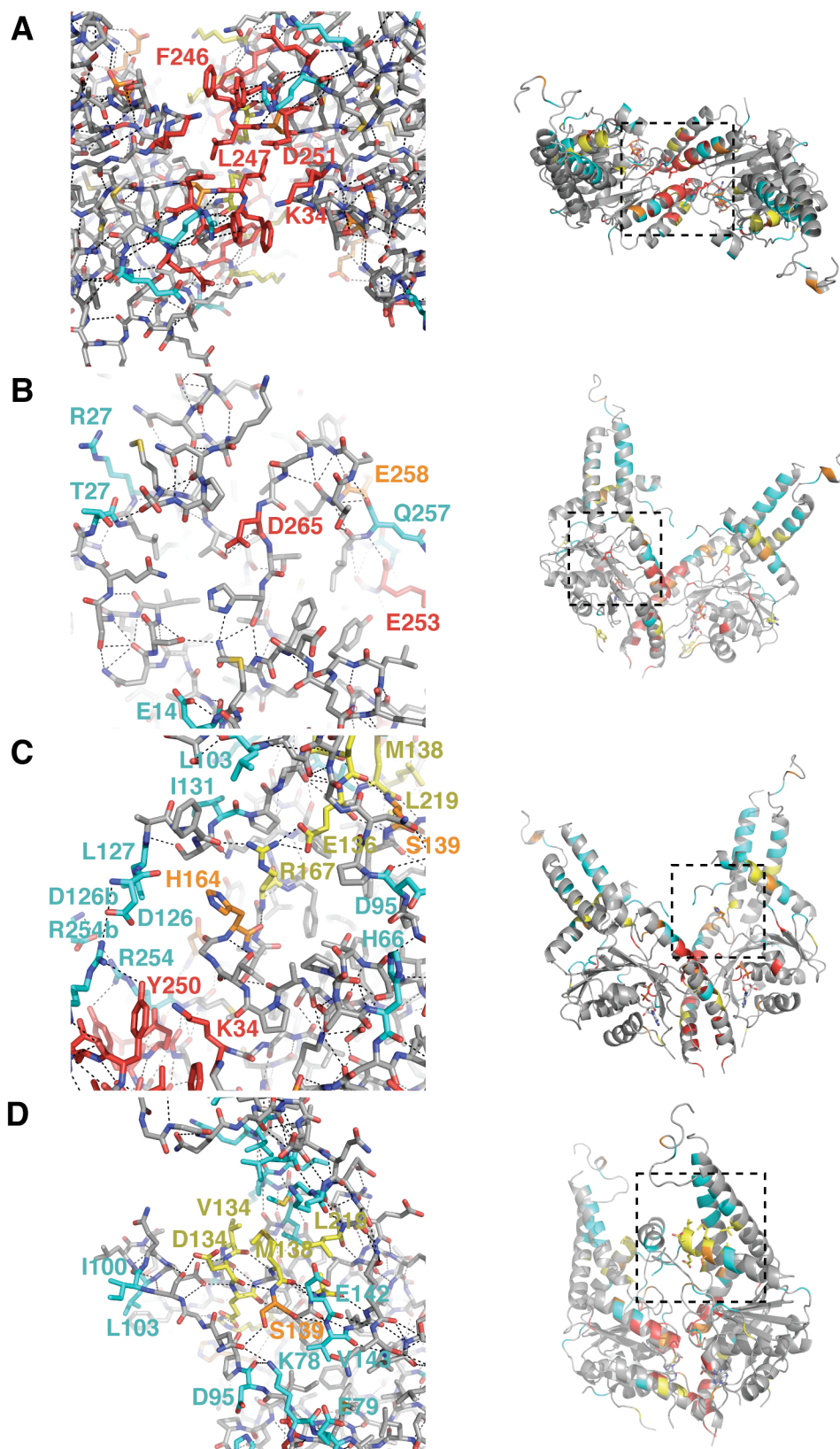


Supplementary Figure 2. The *Sr*Get3 (A–C) and *Aj*Get3 (D–F) crystal forms. (A) The *Sr*Get3-apo monomer with labeling and color similar to Fig. 1 A. The modeled zinc (slate) and putative Ni (green) are shown as spheres with coordinating side-chains from the monomer as sticks. (B) A composite omit map contoured at 1σ with the protein model in sticks colored as in A. (C) Opposing crystallographic *Sr*Get3-apo dimer showing the oriented SB1/2 loops that interact in the crystal lattice. One monomer is colored by motifs and the other is in salmon. The SB1/2 loops are generally disordered and not clearly interpretable in our structure. (D) A monomer of the *Aj*Get3-apo dimer found by molecular replacement. Only the portions of the *Aj*Get3-ADP structure used as a search model are shown. Density is a 1.2σ map calculated using phases from the molecular replacement solution. Additional density for the truncated SB2 can be seen. (E) The NHD dimer interface is slightly different in the apo form. We have modeled this movement using the *Aj*Get3-ADP form. That dimer is in gray, and the movement in the apo form is modeled in purple. (F) The arm dimer seen in the crystal packing for the apo form. Loops are poorly ordered and the ADP form is modeled as transparent helices for clarity.

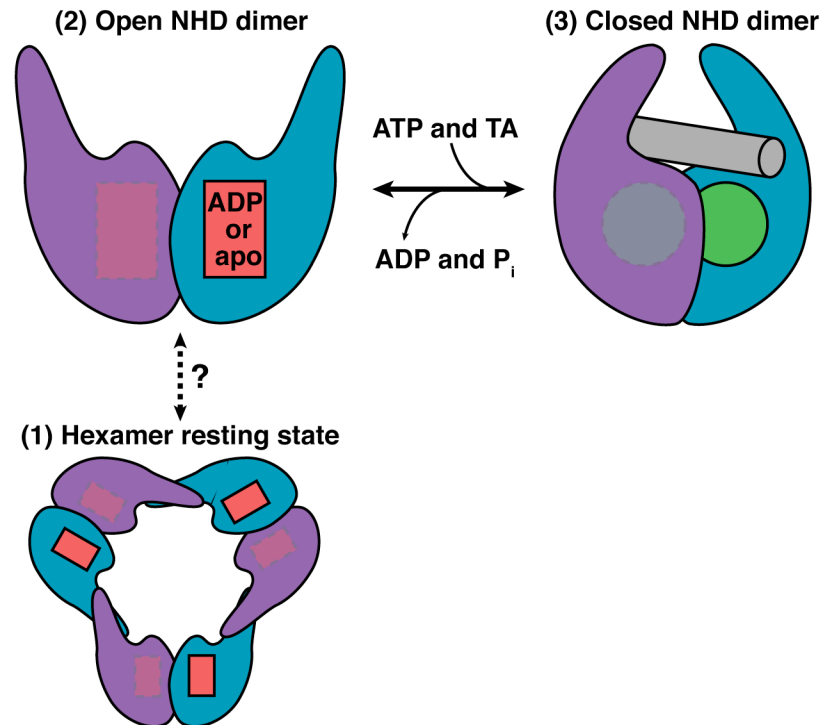




Supplementary Figure 3. Mutant rescue experiments. The full panel of yeast mutants similar to Fig. 4 A. Subpanel letters indicates a group of mutants plated together, one plate for each growth condition. Every plate contained the parent strain, knockout strain and plasmid complemented transformant as functional controls. Growth defects were tested on SC-Ura supplemented with 2 mM CuSO₄ at 30° C and 37° C, 200 mM hydroxyurea at 30° C, and SC-Ura at 40° C. SC-Ura at 30° C was used as a growth condition control. Experiments not performed are highlighted by hashed rectangles. Our interpretation of these results is indicated in Table S1.



Supplemental Figure 4. Phenotype regions described in the text. (A–D) On the left are zoomed in views of the regions described in the text. Residues are shown as sticks and colored as in 4B and some of the phenotype residues are labeled. Hydrogen-bonds are shown as gray dashes. On the left is a ribbon diagram of the full dimer in the same orientation with a few of the residues drawn as sticks for orientation. The dashed boxes indicated the approximate region shown on the right.



Supplemental Figure 5. Model for tail-anchored recognition. The discussion suggests the following model. In a resting/open form Get3 may alternate between (1) a hexamer and (2) a NHD dimer. Binding of ATP would lead to a conformational change that would facilitate formation of a stable complex with a TA protein. Binding of both would result in a complex (3) primed for ATP hydrolysis. Additional conformational changes or partner binding would stimulate ATP hydrolysis. Loss of the inorganic phosphate (P_i) would lead to release of the TA protein and a return to the open form.

	<i>Aj</i> Get3-ADP	<i>Aj</i> Get3-ADP MAD			ScGet3-apo	<i>Aj</i> Get3-apo
Data collection						
Space group	P2 ₁ 2 ₁ 2 ₁	P2 ₁ 2 ₁ 2 ₁			H32	P4 ₂ 32
Cell dimensions						
<i>a</i> , <i>b</i> , <i>c</i> (Å)	67.76, 154.78, 242.88	68.72, 155.51, 242.72			115.32, 115.32, 281.11	181.02, 181.02, 181.02
α , β , γ (°)	90, 90, 90	90, 90, 90			90, 90, 120	90, 90, 90
Resolution (Å)	50-3.2(3.37-3.2)	50-4.5(4.74-4.5)			50-3.7(3.9-3.7)	50-7.5(7.91-7.5)
		<i>Peak</i>	<i>Inflection</i>	<i>Remote</i>		
Wavelength (Å)	1.00000	0.97941	0.97954	0.91837	1.00000	1.00462
<i>R</i> _{merge} (%)	12.4(66.7)	8.0(13.4)	7.7(12.0)	7.7(12.5)	9.9(62.7)	6.4(41.3)
<i>I</i> / σ <i>I</i>	10.4(2.5)	18.0(12.7)	19(13.5)	18.6(13.5)	8.9(2.7)	16.9(3.9)
Completeness (%)	100.0(100.0)	100.0(100.0)	100.0(100.0)	100.0(100.0)	99.9(100.0)	97.1(97.9)
Redundancy	4.9(5.0)	7.3(7.5)	7.3(7.5)	7.3(7.5)	5.9(6.0)	5.0(5.2)
Refinement						
Resolution (Å)	50-3.2				50-3.7	
No. reflections	43100				7936	
<i>R</i> _{work} / <i>R</i> _{free} (%)	21.2/25.13				28.1/33.5	
No. atoms						
Protein	14082				2184	
Ligand/ion	156				2	
<i>B</i> -factors						
Protein	86				182	
Ligands/Ions	71				120	
Bond RMSD						
Lengths (Å)	0.01				0.007	
Angles (°)	1.362				1.147	

High-resolution shell in parenthesis.

Supplementary Table 1. Crystallographic statistics

Numbering		Phenotypes					Figure S3 Plate set
<i>Δ</i> Get3	<i>A</i> / <i>Get3</i>	2mM Cu 30°C	2mM Cu 37°C	200mM HU 30°C	40°C	overall	
E6A	E14	W		W			C, D
T17A	T25	W	W		W		E, I
K19A	R27						E, I
K26A	K34	S	S	S	S	S	F, I
K26R	K34	M	M	M	S	S	F, I
G30R	G38	S	S	S	S	S	C, D
H60A	H66	W		+			F, I
D64S	D70	S	S	W	M	S	E, I
K69A	K75	+		W	M	+ / M	C, D
K72A	K78	W					F, I
D73A	D79	W					F, I
R75A	R81	W	W	W	W	W	A, D
E87A	E93	S	M	W	W	S	E, I
D89A	D95						E, I
V102S	I100		W				A, D
A105S	L103	W		W			F, I
L117S	L115	W	W	W		W	F, I
L120S	L118						F, I
G123P	G121						F, I
A125S	M123		W	W	W	W	A, D
L126S	M124						C, D
D128A	D126			W	W		A, D
L129S	L127						F, I
I133S	I131				W		F, I
I136S	V134	W		W	W	W	A, D
I136D	V134	W			W	W	F, I
D137A	D135	W		W	W	W	A, D
E138A	E136	W	W		W	W	E, I
L140S	M138	W		W	M	W	A, D
S141A	S139	W	W	W	M	M	A, D
M143S	A141	W	W	W	W	W	A, D
E144S	E142				W		C, D
V145S	V143						F, I
H172A	H164	M	M	W	W	M	F, I
R175A	R167	M	W	W	W	W	E, I
L183S	L175						F, J
K185A	K177						B, D
L186S	A178						F, J
L187S	L179						C, D
K189A	K181						B, D
F190S	L182						F, J
I193S	L185						F, J
M200S	M192						B, D
L201S	L193		W		W		B, E

Numbering		Phenotypes					Figure S3 Plate set
<i>Δ</i> Get3	<i>A</i> /Get3	2mM Cu 30°C	2mM Cu 37°C	200mM HU 30°C	40°C	overall	
N202A	N194						B, E
G206P	G202						F, J
I212S	L212						F, J
K215A	K215						B, E
L216S	M216						F, J
L219S	L219	W	M			W	G, J
E245A	R245		W		W	W	E, J
F246A	F246	S	S	S	S	S	E, J
L247S	L247	M	M	M	S	S	G, J
S248A	S248	W	M	W	S	M	G, J
Y250A	Y250	S	S	S	S	S	G, J
E251A	E251	S	S	S	S	S	G, J
E253A	E253	S	S	S	S	S	G, J
R254A	R254				W		E, J
Q257A	Q257	W	W		W		E, J
E258A	E258	W	M		M	M	G, J
D265A	D265	S	S	M	S	S	B, E
C285T- C288T	C283T- C286T	S	S	S	S	S	C, D
R291A	R289	S	S	S	S	S	B, E
M294A	M292	W	W		W	W	E, J
K297A	K295	W	W	W	W	W	E, J
Y298A	Y296	M	M	S	S	S	E, J
D300A	E298						G, H
E320A	E318	W	M		M	M	E, H
Y338A	Y336	W	W	W	W	W	B, E
G30	G38R	M	S	M	M	M	C, D
N/A	R200A		M		M	M	G, H

Phenotypes in bold represent mutants that had been tested in previous studies. Mutants that could rescue knockout are unmarked. Mutants that could not fully rescue are shown by minus signs are graded by severity of growth defect: weak (W), moderate (M) and strong (S). Mutants that showed apparent gain of function are shown by plus signs (+).

Supplementary Table 2. Summary of mutants

Chapter 2

TOWARD CRYSTALLIZATION OF A GET3 AND TAIL-ANCHORED PROTEIN COMPLEX

Introduction

Membrane protein localization is an essential process in the cell, and its importance is reflected in the universal conservation of the signal recognition particle (SRP) pathway for protein translocation (reviewed in Keenan et al., 2001; Shan and Walter, 2005). An important class of proteins is not accessible to the SRP pathway, however, due to their architecture. These tail-anchored (TA) proteins have a single transmembrane domain close to the C-terminus preventing co-translational delivery (Kutay et al., 1993). TA proteins perform a diverse and critical set of functions as enzymes, protein transport and vesicle trafficking factors, apoptosis regulators and viral components (Beilharz et al., 2003; Kalbfleisch et al., 2007). The GET (Guided Entry of Tail-anchored proteins) pathway has been shown to post-translationally target TA proteins to the endoplasmic reticulum (ER) membrane (reviewed in Hegde and Keenan, 2011).

The first component of the GET pathway discovered was Get3 (TRC40 in mammals) and was shown to bind specifically to the transmembrane domain (TMD) of TA proteins (Stefanovic and Hegde, 2007; Favaloro et al., 2008). Next, a receptor complex of Get1 (WRB in mammals) and Get2 for Get3 in the ER membrane was found (Schuldiner et al., 2008; Vilardi et al., 2011). This was followed by the discovery of a complex for sorting and loading TA proteins into Get3, Get4/5 and Sgt2 in *S. cerevisiae* and the Bag6 complex in mammals (Jonikas et al., 2009; Chang et al., 2010; Costanzo et al., 2010; Leznicki et al., 2010; Mariappan et al., 2010).

Get3 is a homodimeric ATPase that functions as the central chaperone for receiving TA proteins from the sorting complex and delivering them to the membrane via the receptor.

Structural studies revealed that an open and closed state of Get3 is regulated by nucleotide (Bozkurt et al., 2009; Hu et al., 2009; Mateja et al., 2009; Suloway et al., 2009; Yamagata et al., 2010). The apo and ADP bound forms are generally in the open state. In the ATP bound or transition state, Get3 undergoes conformational changes to form the closed state with movement in the relative orientation of the two subunits and rearrangement of helices distal to the nucleotide binding domain (Bozkurt et al., 2009; Mateja et al., 2009; Suloway et al., 2009). Both of these movements cause a binding groove lined with conserved hydrophobic residues to form, which could accommodate the TMD of TA proteins (Bozkurt et al., 2009; Mateja et al., 2009). Mutagenesis of the hydrophobic binding groove by introducing polar or charged residues showed decreased binding of TA proteins to Get3 and reduced rescue of growth defects in *S. cerevisiae* Get3 knockouts (Mateja et al., 2009; Suloway et al., 2009). Disordered regions in the closed form crystal structures extend out from the helices vertically flanking the sides of the binding groove and have been proposed to shield the exposed portion of the binding groove. These regions are ordered in some open form structures through crystal contacts and have been shown to be essential for TA protein binding (Yamagata et al., 2010). They may play a more complex role than initially proposed.

The TMD of TA proteins varies in length and composition, and how Get3 accommodates or selects for these differences is a fundamental question about the function of the GET pathway. It has been suggested that positive charges on either end of the binding groove could reduce the propensity to bind longer mitochondrial TAs that are flanked by positive charge at the C-terminus (Borgese et al., 2007). How conformational change in Get3 induced by

nucleotide state and hydrolysis acts on the release of TA proteins is also a question. The structure of Get3 complexed with TA proteins would reveal the molecular mechanism for interaction of Get3 with TA proteins, give insight into the function of the GET pathway and be the basis for future studies.

The following chapter documents two techniques for forming a complex of Get3 with TA proteins for structural studies. First, recombinant co-expression of Get3 and TA proteins in *E. coli* produces milligram amounts of complex that can be purified. Second, in vitro reconstitution of the complex by combining solubilized TA protein with Get3 allows the addition of nucleotide and cofactors for formation of a complex in a controlled state. Structure determination by X-ray crystallography of recombinant complex was attempted with crystallization trials, crystal refinement and preliminary diffraction data to 11–12Å.

Results

Cloning and Co-expression of Get3 and TA Proteins

For this study our ultimate goal was to generate sufficient amounts of homogenous complex of Get3 with a TA protein for crystallographic structure studies. Generally, to efficiently achieve sufficient amounts of proteins for structural studies recombinant *E. coli* overexpression systems are used. TMDs are composed of a hydrophobic stretch of amino acids that normally reside in the membrane and are not part of the folded soluble portion of the protein. They are unprotected from the polar solvent unless shielded by chaperones until integrated into the membrane. Recombinant expression of TA proteins in *E. coli* would result in membrane integration, cytosolic aggregation or degradation. Overexpression could lead to overwhelming endogenous pathways for membrane integration leading to degradation or toxicity. Furthermore, TA proteins would have to be solubilized from the membrane or aggregates to form complex with Get3.

To avoid these complications and directly generate complex, Get3 was recombinantly co-expressed in *E. coli* with the TA protein. Recombinant Get3 expressed in the cell would bind to TA proteins directly after synthesis or obtain them from endogenous chaperone mechanisms. In the absence of the Get1/2 receptor, complexes would remain intact for extraction and purification. To facilitate expression of different combinations of TA proteins and Get3 constructs without additional cloning, TA proteins were cloned into a separate vector from Get3 constructs so the two plasmids could be co-transformed.

Tail-anchored proteins from previous bioinformatics surveys (Beilharz et al., 2003; Kalbfleisch et al., 2007) were selected for cloning to give a diversity in the N-terminal cytosolic domain, TA and C-terminal ER resident region. The region N-terminal to the TMD is longer than the

C-terminal region, which is often less than 10 amino acids, so tags for affinity purification were placed at the N-terminus of the TA-proteins to avoid disruption of TMD dependent complex formation and to ensure accessibility of the tag. *S. cerevisiae* and *H. sapiens* TA proteins were N-terminally His-tagged in the first MCS of the pACYCDuet vector. An N-terminal MBP fusion to the TA protein Sbh1 separated by a thrombin protease cleavage site was cloned into the pMAL-C2 vector. The TA protein Sec61 β was cloned into pET33b with an N-terminal FLAG tag. Table 1 provides a list of cloned TA protein constructs.

For co-expression of Get3 and TA proteins an MBP fusion to Get3 in pMAL-C2 or an untagged Get3 construct in pET33b was co-transformed with His-tagged TA proteins in pACYCDuet into the *E. coli* expression strain BL21Gold. Co-transformations with MBP tagged Sbh1 in pMAL-C2 were carried out with a His-tagged Get3 or untagged Get3 in pACYCDuet. BL21Gold co-transformants were selected using the antibiotic resistant markers from both plasmids, grown at 37 C and induced for 3–4 h before harvesting. Expression of both Get3 and the TA proteins was typically strong enough to be seen by comparing the pre- and post-induction samples using SDS-PAGE (Supplemental Figure 1A).

Purification of Get3 Complexed to TA Proteins

Complex purification from recombinant co-expression relies on the TA protein binding stably to Get3 in *E. coli* so that it can be extracted. Cultures induced for co-expression were pelleted, resuspended in lysis buffer and lysed. Cellular debris, including protein aggregates, were then pelleted and separated from the supernatant to yield a clarified extract. Stable complex formation present in the cell extract would remain soluble after centrifugation, and the presence of both the TA protein and Get3 could be confirmed by SDS-PAGE of the clarified extract (Figure 1A & B).

To verify that TA protein was bound to Get3 instead of it being soluble on its own or chaperoned by an *E. coli* protein, a two-step affinity purification was performed using different affinity tags on the TA protein and Get3. In the first step of affinity purification of Get3 TA protein was present in elution fractions. For the second step affinity purification of the TA protein from the first step elution, unbound Get3 was washed away revealing only complex of Get3 and TA protein (Figure 1A). A one-step purification was also possible by affinity purification of only the TA protein as TA protein not bound to Get3 was insoluble, yielding only complex (Figure 1B).

Size-exclusion chromatography (SEC) was used for the final purification step for the complex and showed Δ Get3 and the TA protein co-eluted in one peak, verifying complex formation (Figure 2). Interestingly, this peak corresponds to a dramatically shorter elution time indicating that even for small TA proteins the complex has a significantly larger hydrodynamic radius compared to Get3 alone. Complexed Get3 could be separated from uncomplexed Get3 as two separate peaks from SEC (Figure 2A). SDS-PAGE showed the first peak contains Get3 and TA protein whereas the second peak only contains Get3 (Figure 2B). Given the small size of TA proteins in the complex, it was likely Get3 exists in a higher oligomeric state than a dimer in the complex when recombinantly co-expressed and purified from *E. coli*.

Crystallization of the Complex

Fractions collected from SEC corresponding to the peak with Get3 complexed with TA protein were used for crystallization trials. Complexes of Δ Get3 with His-tagged Sbh1, Sbh2 or Ysy6 were concentrated to 10–15mg/ml and dialyzed in crystallization buffer before use in crystallization screens. For initial screening commercial screens including the Crystal HT, Index HT, MembFac, PEG/Ion Screen, Wizard, PACT *premier*, and JCSG+ Suite were used.

Most crystals appeared in screens for *Sa*Get3 in complex with Sbh1 with the addition of ADP·AlF_x, totaling 11 unique conditions. Crystals also appeared in one condition with the non-hydrolyzable ATP analog AMPPNP. Two conditions grew crystals for *Sa*Get3 in complex with Sbh2 and one condition for complex with Ysy6, all with ADP·AlF_x. Hits for Index H4, Wizard D6, PACT E1, Crystal D4, MembFac A3, Wizard C3 and PACT E11 were refined with a gradient around each of the original conditions (Supplemental Figure 2).

In order to verify the crystals contained complex, they were subjected to SDS-PAGE. Three of the conditions with the largest single crystals (Figure 3A-C) had sufficient material for analysis and were removed from the drops, washed three times in the crystallization condition and resuspended in SDS-PAGE loading buffer. The samples were run on an SDS-PAGE gel and silver stained to detect Get3 and the TA-protein. Clear bands were visible at the molecular weight expected for Get3, but no TA protein bands were visible (Figure 3D). It is possible TA proteins are present but not in sufficient quantity (sub-microgram amounts of protein from the crystals) to be visible. The small size and hydrophobic nature of the TA proteins could also result in poor staining results.

Since the presence of complex in the crystals was unclear, X-ray diffraction data were collected using the crystals in hopes that the structure would reveal if TA protein was bound to Get3. Of the crystals that diffracted, the highest resolution diffraction spots were at 11Å for *Sa*Get3 complexed with Ysy6 with MgADP·AlF_x from the refined crystallization condition 0.1M Sodium citrate and 20% PEG 3350, cryoprotected with 20% glycerol (Figure 3E). The best diffracting crystal from *Sa*Get3 complexed with Sbh1 was in the same buffer as the Ysy6 complex and was from the refined crystallization condition 0.1M Sodium citrate pH 6.2, 10% 2-propanol and 20% PEG 4000. This crystal diffracted to 12Å when cryoprotected with 20%

glycerol. Due to the low resolution of the diffraction pattern no definitive indexing solutions could be found with iMosflm (Battye et al., 2011) for either crystal.

Alternative complexes of Get3 with TA proteins were tested. Co-expression of *Hs*Get3/TRC40 or *Aj*Get3 with TA proteins failed to give a clean SEC profile. In an attempt to purify the complex in an ATP bound form, potentially ATPase inactive mutants of *Sa*Get3 G27P, G28P, K31A,/R, D57I/K/L/V were co-expressed with Sbh1 and purified. G27P and D57L were selected for SEC, but G27P formed mostly aggregates and eluted in the void volume and D57L formed only small amounts of complex. Other TA protein substrates with larger and more complex N-terminal domains were tested but failed to elute in separable peaks (Table 2).

In vitro Complex Reconstitution

Crystals from conditions with *Sa*Get3 complexed with TA proteins formed recombinantly failed to diffract to sufficient resolution for structure determination, and the presence of TA protein in the crystal could not be confirmed by SDS-PAGE. Since nucleotide state had been shown to be important for Get3 conformation (Bozkurt et al., 2009; Mateja et al., 2009), the presence of nucleotide was checked by measuring the magnesium concentration within the recombinant complex by inductively coupled plasma mass spectrometry (ICP-MS) (Yamagata et al., 2010). Although the concentration of zinc matched what would be expected to be coordinated by the CxxC motif of each dimer, the concentration of magnesium was approximately sixfold less (Table 3). Without the presence of magnesium it was unlikely there was nucleotide present in the complex. In previous crystallization trials magnesium and nucleotide analogs were added, but since bound TA protein could serve to stabilize the closed conformation of Get3, the nucleotide state was uncertain. Combining Get3 and the TA

protein purified separately in vitro would allow the addition of nucleotide in a manner already shown to bind Get3 and possibly make a more stable and physiologically relevant complex.

Purifying the TA protein when expressed without Get3 requires it be made soluble by other means. One standard technique for solubilizing proteins is to use the denaturants urea or guanidine. The TA protein Ysy6 with an N-terminal His-tag was expressed in the *E. coli* expression strain BL21Gold. After centrifugation the pellet was resuspended in either urea or guanidine to lyse the cells and solubilize Ysy6, which was then affinity purified with Ni-NTA resin. Although concentration of urea from 1 to 8M solubilized Ysy6, it could be pelleted by ultracentrifugation indicating it was forming large, partially soluble aggregates (Figure 4A). When 6M guanidine was used, Ysy6 remained in the supernatant after ultracentrifugation and thus accessible for complex formation (Figure 4B).

Using TA protein solubilized in denaturant to form complex without denaturing Get3 would eliminate the need to refold the entire complex bound to nucleotide at once. One technique to accomplish this would be to rapidly dilute the denatured TA protein into a buffer containing Get3, analogous to refolding proteins from denaturants. Ysy6 denatured in 6M guanidine was concentrated to $\sim 175\mu\text{M}$ in $150\mu\text{l}$ and rapidly diluted into 15ml of $\sim 30\mu\text{M}$ Get3 in TBS with either ADP, ATP, ADP $\cdot\text{AlF}_x$ or without added nucleotide. This corresponds to a slight molar excess of Ysy6 to a Get3 dimer. After dilution the sample was ultracentrifuged to remove aggregates and added to NiNTA resin and then washed. Get3 complexed with Ysy6 was eluted with imidazole (Figure 4C). Elution fractions show complex formation in ATP and ADP $\cdot\text{AlF}_x$ containing buffers but not for ADP. Surprisingly, the most complex was formed without nucleotide. The yield of complex was low with only microgram amounts of complex formed from more than 10mg of Get3. SEC was attempted for complex samples, but they

were not confidently detected by 280 or 214nm absorbance. Achieving milligram amounts of complex for crystallization trials by this would require further optimization of the protocol to reduce the amount of purified Get3 needed.

Figures

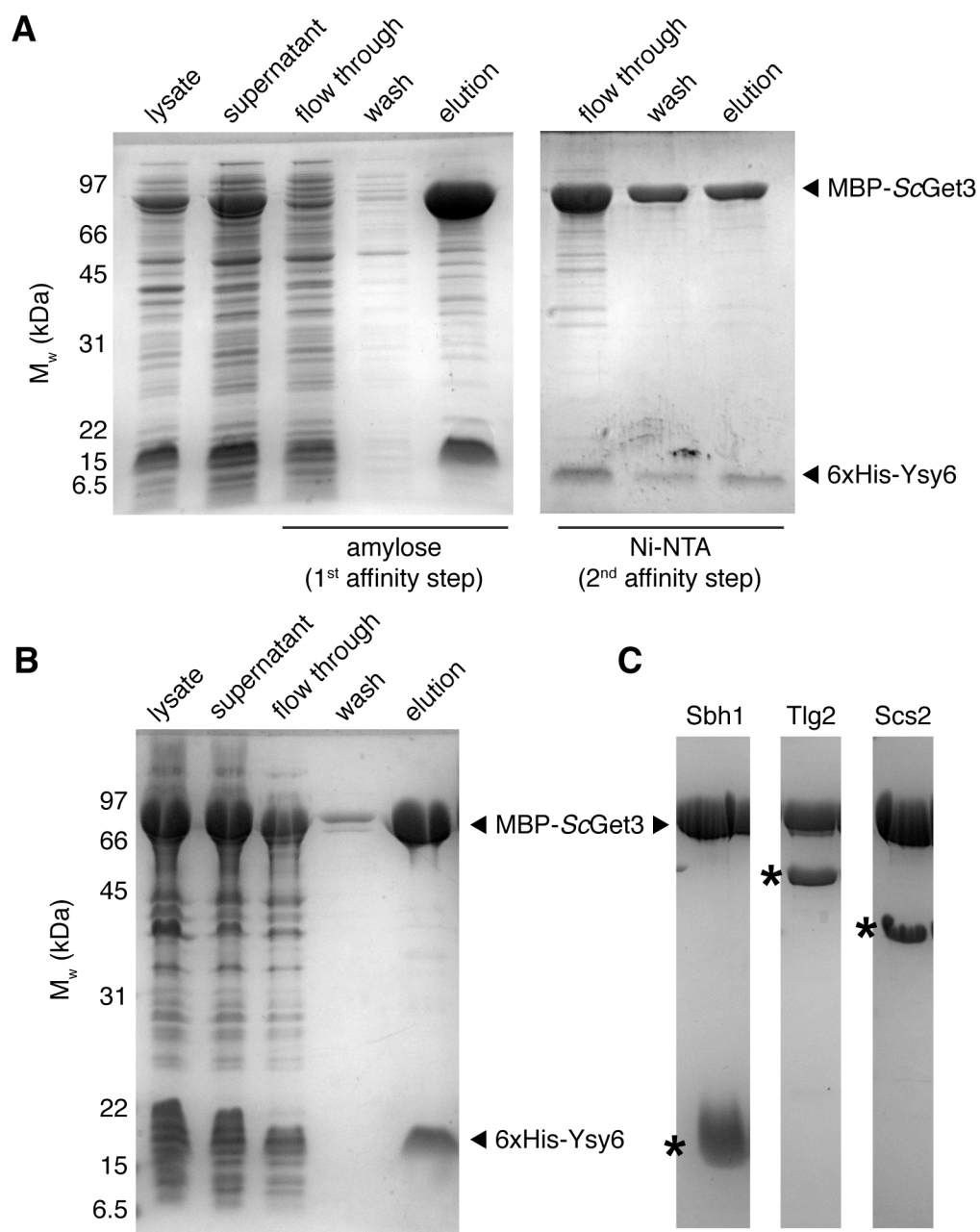


Figure 1. Affinity purification of recombinant co-expression of *Ss*Get3 complexed with TA proteins. (A) Two-step complex affinity purification analyzed by coomassie-stained SDS-PAGE. For the left gel the first two lanes are lysate and then supernatant after centrifugation. The next three lanes are the flow through, wash and eluate from the first step of affinity purification using amylose resin for the MBP fusion to Get3. The final three lanes in the gel on the right are the flow through, wash and eluate from the second step of affinity purification using Ni-NTA resin for the His-tagged TA protein Ysy6. Bands for MBP-*Ss*Get3 and 6xHis-Ysy6 are labelled. (B) One-step complex affinity purification analyzed by coomassie-stained SDS-PAGE. Similar to the left gel of (A) except the affinity purification used Ni-NTA resin for the His-tagged TA protein in complex with MBP-*Ss*Get3. (C) Coomassie-stained SDS-PAGE purified complex of MBP-*Ss*Get3 with three additional TA proteins: Sbh1, Tlg2 and Scs2. MBP-*Ss*Get3 is labelled and asterisks indicate bands corresponding to the TA proteins.

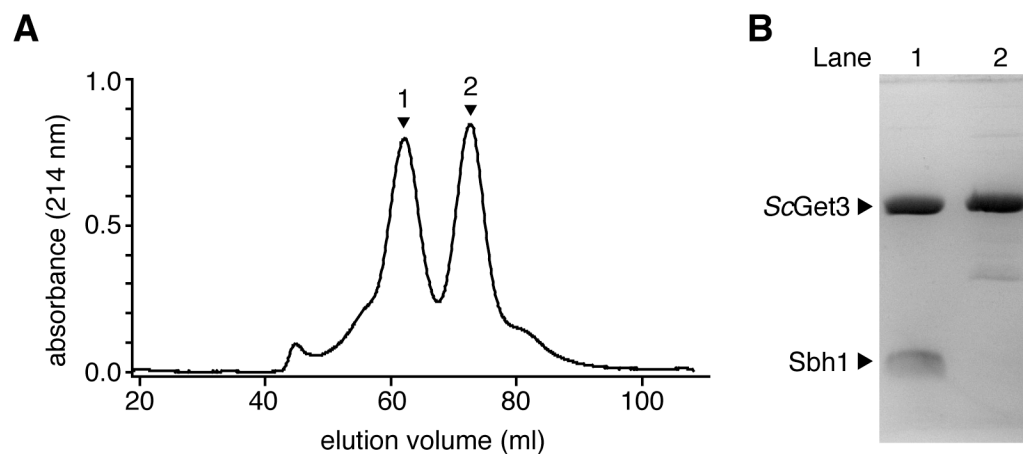


Figure 2. Size exclusion chromatography of *ScGet3* complexed with TA protein. (A) A size-exclusion chromatograph of *ScGet3* complexed with Sbh1 with the two prominent peaks labelled 1 and 2. (B) Coomassie stained SDS-PAGE of SEC fractions with lanes 1 and 2 corresponding to peaks 1 and 2 in (A). The faster eluting peak 1 has a bands for both *ScGet3* and Sbh1 (labelled) whereas peak 2 has a band for *ScGet3* but not Sbh1.

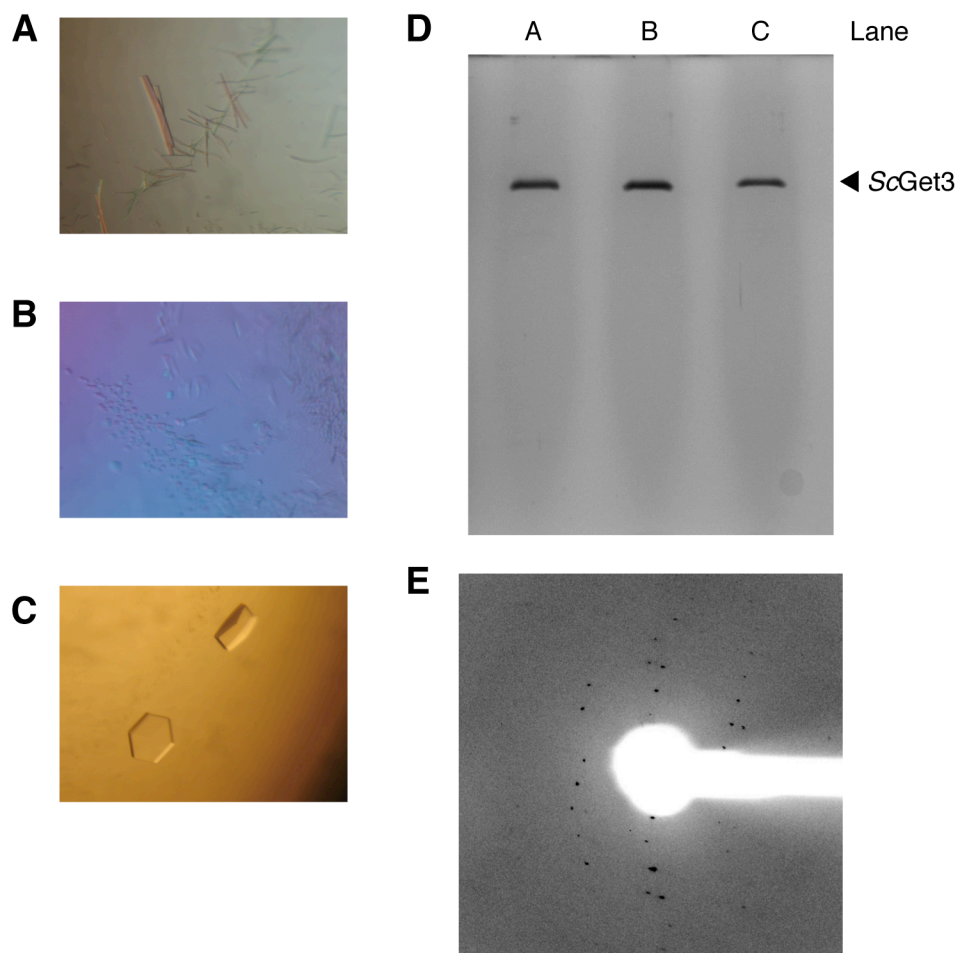


Figure 3. Crystals from conditions with Get3 complexed with TA proteins. (A) *ScGet3* complexed with Sbh1 in 0.2M ammonium citrate pH 7.0 and 20% PEG 3350. (B) *ScGet3* complexed with Sbh1 in 0.1M sodium citrate pH 5.6, 20% 2-propanol and 20% PEG 4000. (C) *ScGet3* complexed with Ysy6 in 0.2M sodium citrate and 20% PEG 3350. (D) Silver-stained SDS-PAGE of washed crystals from crystallization conditions with *ScGet3* complexed with TA proteins. Lanes 1–3 are from crystals in A–C. The arrow indicates bands corresponding to *ScGet3*. (E) Diffraction image from *ScGet3* complexed with Ysy6 with MgADP·AlF_x in 0.1M Sodium citrate and 20% PEG 3350, cryoprotected with 20% glycerol. The edges of the image corresponds to 11.5Å diffraction in the shortest dimension.

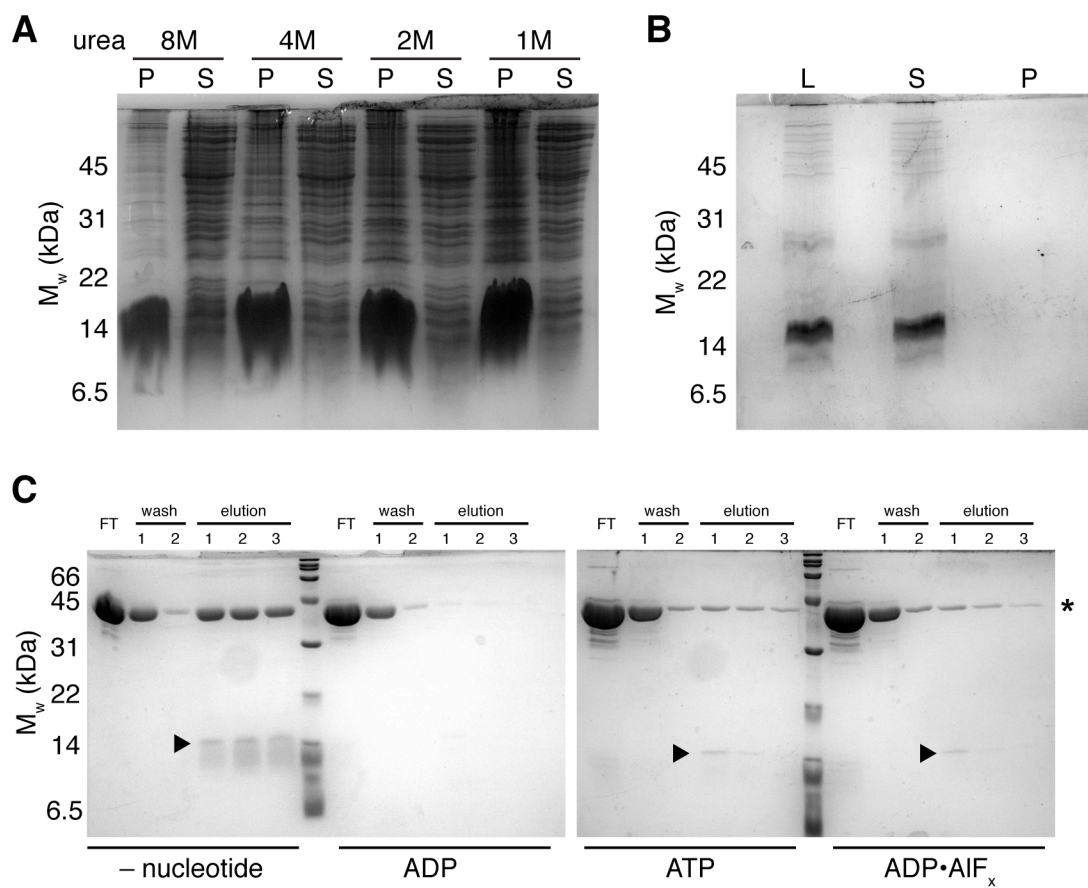


Figure 4. **In vitro reconstitution of Get3 complexed with TA protein.** (A) Ysy6 affinity purified under denaturing conditions using urea. Stability was tested by ultracentrifugation with the pellet (lanes “P”) and supernatant (lanes “S”) analyzed by coomassie-stained SDS-PAGE. In 8, 4, 2 and 1M urea Ysy6 was unstable and pelleted. (A) Ysy6 affinity purified under denaturing conditions using guanidine. As for (A) except the load (L) is shown for comparison to the supernatant (S) and Ysy6 remained stable in 6M guanidine with no visible band in the pellet (P). (C) Purification of in vitro reconstituted complex in different nucleotide states. Coomassie stained SDS-PAGE of apo, ADP, ATP and ADP·AlF_x in vitro reconstitutions with fractions from the Ni-NTA affinity purification flow through, washes 1 & 2 and elutions 1–3. Get3 pulled down (band marked by an asterisk) as a complex with His-tagged Ysy6 (bands marked by arrowheads) in the elution fractions from apo, ATP and ADP·AlF_x containing buffers.

Protein	Organism	Construct	Vector
Sec61 β	<i>H. sapiens</i>	6xHis-Sec61 β (Δ TMD)	pACYCDuet
Sec61 β	<i>H. sapiens</i>	6xHis-Sec61 β	pACYCDuet
Ysy6	<i>S. cerevisiae</i>	6xHis-Thr-Ysy6 (residues 37-65)	pACYCDuet
Ysy6	<i>S. cerevisiae</i>	6xHis-Ysy6	pACYCDuet
RAMP4/SERP1	<i>H. sapiens</i>	6xHis-SERP1	pACYCDuet
Pex15	<i>S. cerevisiae</i>	6xHis-Pex15	pACYCDuet
Tlg2	<i>S. cerevisiae</i>	6xHis-Tlg2	pACYCDuet
Dpm1	<i>S. cerevisiae</i>	6xHis-TEV-Dpm1	pACYCDuet
Fis1	<i>S. cerevisiae</i>	6xHis-Fis1	pACYCDuet
Sed5	<i>S. cerevisiae</i>	6xHis-Sed5	pACYCDuet
Cyb5	<i>H. sapiens</i>	6xHis-Cyb5	pACYCDuet
Sbh1	<i>S. cerevisiae</i>	6xHis-Sbh1	pACYCDuet
Sbh1	<i>S. cerevisiae</i>	6xHis-6xArg-Sbh1	pACYCDuet
Sbh2	<i>S. cerevisiae</i>	6xHis-Sbh2	pACYCDuet
Scs2	<i>S. cerevisiae</i>	6xHis-Scs2	pACYCDuet
Mga2	<i>S. cerevisiae</i>	6xHis-Mga2	pACYCDuet
Mga2	<i>S. cerevisiae</i>	6xHis-Mga2 (Δ TMD)	pACYCDuet
Sec22?	<i>S. cerevisiae</i>	6xHis-Sec22	pACYCDuet
Sbh1	<i>S. cerevisiae</i>	MBP-Thr-Sbh1	pMAL-C2
Sbh1	<i>S. cerevisiae</i>	MBP-Thr-Sbh1 (residues 52-82)	pMAL-C2
Sbh1	<i>S. cerevisiae</i>	MBP-Thr-Sbh1 (residues 47-82)	pMAL-C2
Sbh1	<i>S. cerevisiae</i>	MBP-Thr-Sbh1 (residues 42-82)	pMAL-C2
Sbh1	<i>S. cerevisiae</i>	MBP-Thr-Sbh1-Thr-MBP (residues 47-82)	pMAL-C2
Sec61 β	<i>H. sapiens</i>	FLAG-Sec61 β (Δ TMD)	pET33b
Sec61 β	<i>H. sapiens</i>	FLAG-Sec61 β	pET33b

Table 1. TA protein constructs cloned for co-expression

Construct	Expressed	Co-expressed	Affinity	SEC	Crystallized
6xHis-Sec61 β (Δ TMD)	+				
6xHis-Sec61 β	+	+	+	–	
6xHis-Thr-Ysy6 (residues 37-65)	+	+	–		
6xHis-Ysy6	+	+	+	+	+
6xHis-SERP1	+				
6xHis-Pex15	+	+	+	–	
6xHis-Tlg2	+	+	+	–	
6xHis-TEV-Dpm1	+	–			
6xHis-Fis1	+	+			
6xHis-Sed5	+				
6xHis-Cyb5	+	+	–		
6xHis-Sbh1	+	+	+	+	+
6xHis-6xArg-Sbh1	+	+	+	+	
6xHis-Sbh2	+	+	+	+	+
6xHis-Scs2	+	+	–		
6xHis-Mga2	+	–			
6xHis-Mga2 (Δ TMD)	+				
6xHis-Sec22	+				
MBP-Thr-Sbh1	+	+	+	+	–
MBP-Thr-Sbh1 (residues 52-82)	+	+	+	+	
MBP-Thr-Sbh1 (residues 47-82)	+	+	+	+	–
MBP-Thr-Sbh1 (residues 42-82)	+	+	+	+	
MBP-Thr-Sbh1-Thr-MBP (residues 47-82)	+				
FLAG-Sec61 β (Δ TMD)	+				
FLAG-Sec61 β	+				

Table 2. Per substrate complex formation and purification with *SrGet3* Constructs correspond to Table 1. Plus signs indicate successful completion of the step, minus signs indicate unsuccessful attempts and blank entries have not been tested. Co-expression was performed with *SrGet3*.

Component	Measured concentration	Molar concentration	Expected concentration
Get3/Sbh1	561 mg/L	6.79 μ M	N/A
Mg(H ₂)	51 μ g/L	1.94 μ M	13.59 μ M
Zn	370 μ g/L	5.66 μ M	6.79 μ M

Table 3. Co-factors in Get3 complex with TA proteins by ICP-MS. Each component has a measured concentration, Get3/Sbh1 by 280nm absorbance and Mg(H₂)/Zn by ICP-MS. The molar concentration is calculated based on the estimated molecular weight. The expected concentration is calculated by the molar concentration of Get3/Sbh1 with two magnesium and one zinc ions per Get3 dimer.

Discussion

Since the identification of Get3/TRC40 (Stefanovic and Hegde, 2007; Favaloro et al., 2008) as a novel targeting factor for TA proteins, a number of structural studies have provided information on the molecular mechanisms for its function (Bozkurt et al., 2009; Hu et al., 2009; Mateja et al., 2009; Suloway et al., 2009; Yamagata et al., 2010). Yet there is still no direct structural evidence for how Get3 interacts with TA proteins and how this is coupled to the targeting process. This work demonstrates two methods for generating a complex of Get3 with TA proteins and preliminary work for X-ray structure determination of the complex. Recombinant co-expression of Get3 with TA proteins in *E. coli* yields complex that can be purified in sufficient quantity for crystallization. In vitro reconstitution is capable of forming complex in the presence of nucleotide and with refinement could be used for structural studies. A number of crystal hits from co-expressed complex yielded low resolution diffraction, and future studies could improve diffraction in order to solve the structure of Get3 in complex with TA protein.

TA proteins perform a range of functions including enzymatic activity, protein transport, vesicle trafficking, apoptosis regulation and viral infectivity. This diversity is reflected in the size and structure of their N-terminal cytosolic domains. The length and composition of the TA and C-terminal region varies as well. There are a number of possible TA proteins and constructs to test for co-crystallization with Get3. A minimal binding construct could reduce disorder and flexibility whereas an ordered N-terminal domain could stabilize lattice formation. Using a range of different TA proteins for complex crystallization improves the chances of finding crystallization hits for structure determination.

In order to facilitate rapid screening of complexes, a co-expression system was employed using two plasmids, one encoding a Get3 construct and another with a TA protein. Get3 was successfully co-expressed with a variety of TA proteins. The purification procedure used different affinity tags on Get3 and the TA proteins and a two-step affinity purification to ensure a resulting homogenous complex. As expected for a complex, purified Get3 and TA protein co-eluted in SEC in a single peak, but it should be noted a sizable peak of uncomplexed Get3 was also observed. These two peaks were easily distinguishable due to the complex eluting much sooner than expected for a Get3 dimer, even with relatively small TA proteins. One explanation is that the complex of Get3 with TA proteins contains a higher oligomeric state of Get3 than a dimer.

Purified Get3 TA complex with Sbh1, Sbh2 and Ysy6 were used in crystallization trials. Nucleotide analogs AMPPNP and ADP·AlF_x were added to stabilize the closed form of the dimer that had been proposed for the TA bound state of Get3. Crystals grew in many unique crystallization conditions with a variety of morphologies. The crystals fluoresced under UV illumination indicating they were protein crystals, unless the additives for ADP·AlF_x fluoresced as well. Crystallization hits were refined and conditions that formed the largest single crystals were selected for further analysis. The crystals were analyzed by SDS-PAGE for the presence of TA protein, but this could not be confirmed. One possibility is that there was insufficient TA protein for detection or there was less detection sensitivity for small hydrophobic proteins. Another explanation would be the dissociation of the complex that could be exacerbated by particular crystallization conditions. Since Get3 readily crystallizes, even small amounts of uncomplexed Get3 could form crystals or there could be bias toward conditions that dissociate the complex. More sensitive detection techniques might reveal the presence of TA proteins in the crystals.

Crystals diffracted X-rays to a resolution of 11–12Å, which was insufficient for structure determination. Disorder caused by the flexible N-terminal regions of the TA-proteins could lower the resolution of diffraction. Sbh1 and Ysy6 have small but not insignificant N-terminal domains, and a minimal TA protein construct for complex formation could reduce the disorder in the crystal. Hydrophobic interactions can be nonspecific and different conformations of complex could arise from the same substrate causing heterogeneity and reducing order in the lattice. Similarly, different stoichiometry of TA proteins bound to Get3 could have slightly different conformations that could cause disorder and result in low resolution diffraction. Many of the more general techniques for crystal improvement could be applied as well (Bergfors, 2003; Heras and Martin, 2005; Derewenda and Vekilov, 2006).

Different methods for complex generation might yield a more crystallographically amenable and physiologically relevant complex than recombinant co-expression in *E. coli*. The apparent lack of nucleotide in the recombinant complex is of concern. Structural studies have shown the presence of nucleotide modulates the conformation between an open and closed form. The closed form presents a hydrophobic groove that is a putative binding site for TA proteins, and this conformation could be necessary for stable complex formation. Generally, protein complexes of soluble proteins can be made in vitro by simply combining the component proteins, which allows for manipulations such as co-factor addition, including nucleotides. TA proteins are hydrophobic making them unstable in solution so accommodations must be made to apply this method to forming complex with Get3. Denaturants were used to solubilize TA proteins and rapid dilution into a buffer containing Get3 resulted in complex formation. Interestingly, the apo, ATP and ADP·AlF_x states formed complex, but the ADP state did not. Since the apo state is open and the ADP·AlF_x state is closed, presumably either state can form complex. The apo state could close upon TA binding. The ATP bound state should be closed

before hydrolysis, but even if ATP was hydrolyzed it has been suggested that P_i does not dissociate immediately from Get3 and could remain in the closed state (Bozkurt et al., 2009). Why Get3 cannot form complex when ADP is bound remains an interesting question. Ultimately the yield from in vitro reconstitution was too low for structural analysis of the complex.

For *S. cerevisiae* the in vivo process of loading TA proteins into Get3 involves many additional protein factors including chaperones, Sgt2, Get4 and Get5 (Wang et al., 2010). This mechanism could serve to sort ER TA proteins appropriate for Get3 from other proteins as well as load TA proteins into Get3 to form a complex competent for insertion. When co-expressing Get3 with TA proteins recombinantly, these factors are absent so it is somewhat surprising the complex forms at all. Recombinant complex is functional as it inserts TA proteins into ER derived microsomes in an in vitro reaction (Favaloro et al., 2008; Bozkurt et al., 2009; Favaloro et al., 2010). Nonetheless, whether recombinant or in vitro reconstituted complex represents the physiological state remains a question, especially considering recombinant complex might represent a higher oligomeric state of Get3 than a dimer.

The structure of Get3 complexed with TA proteins would give much-needed insight into the molecular mechanisms of the GET pathway. Two methods for generating a complex of Get3 with TA proteins have been presented here. Recombinant co-expression of Get3 with a TA protein in *E. coli* is capable of forming a complex with a number of different TA proteins. In vitro reconstitution of complex from purified Get3 and denatured TA proteins allows for complex formation in different nucleotide states. Crystals formed from crystallization trials of the complex and could be refined, although the presence of TA proteins is unconfirmed. More work is needed to improve the diffraction of these crystals to be able to determine the

structure with X-ray crystallography. How Get3 interacts with TA proteins and what role the mechanism plays in the GET pathway are important questions about the process of protein translocation.

Materials and Methods

Cloning

S. cerevisiae TA proteins were amplified from S288c (ATCC) genomic DNA except for Ysy6 constructs which were synthesized from primers designed with DNAWorks (Hoover and Lubkowski, 2002). *H. sapiens* TA proteins were all primer synthesized using DNAworks. N-terminally 6xHis-tagged TA proteins Ysy6, SERP1, Fis1, Sed5, Cyb5, Sbh1, Sbh2, MGA2, MGA2 residues 2-1036 and Sec22 were cloned into the first MCS of pACYCDuet (Novagen) using restriction sites for BamHI and PstI. Pex15, Tlg2 and Scs2 were cloned using HindIII instead of PstI due to an internal PstI site. Ysy6 residues 37-65 were synthesized with an N-terminal 6xHis-tag and thrombin protease cleavage site and were cloned into pACYCDuet with NcoI and PstI restriction sites. N-terminally 6xHis-tagged Dpm1 with a TEV protease restriction site was subcloned from a previous construct in pET33b (Novagen) using NcoI and NotI restriction sites. N-terminally FLAG tagged Sec61 β and Sec61 β residues 2–69 were primer synthesized using DNAWorks and cloned into pET33b using NcoI and BglII restriction sites. Both constructs were subcloned into pACYCDuet into SalI and NotI restriction sites to for N-terminally 6xHis-tagged constructs. N-terminal MBP fusions to Sbh1 with a thrombin protease cleavage site in pMAL-C2 (NEB) were created by first using site-directed mutagenesis to change the factor Xa cleavage to a thrombin cleavage site and then cloning PCR amplified Sbh1 into BamHI and PstI sites to account for a frame shift from the pACYCDuet vector BamHI site. N-terminally 6xHis-6xArg-tagged Sbh1 was made by site-directed mutagenesis (Stratagene) of the 6xHis-tagged Sbh1 pACYCDuet construct. Truncations of Sbh1 with an N-terminal MBP fusion in pMAL-C2 were also made by site-directed mutagenesis.

N-terminally 6xHis-tagged *Ss*Get3 and *Aj*Get3 with a thrombin protease cleavage site in pET33b from a previous study were used (Suloway et al., 2009). *Hs*Get3/TRC40 was amplified from the cDNA clone in pCMV-XL6 (OriGene) and cloned in pET33b using SalI and BglII restrictions to make a thrombin cleavable N-terminally 6xHis-tagged construct. Untagged *Ss*Get3 in pACYCDuet was amplified from the 6xHis-Thr-*Ss*Get3 pET33b construct and cloned using NcoI and XhoI restriction sites. Untagged *S. cerevisiae* Get3 in pET33b was subcloned from pACYCDuet using the same restriction sites. N-terminal MBP fusions to *Ss*Get3 and *Aj*Get3 with a thrombin protease cleavage site in pMAL-C2 were cloned into the MBP-Thr-Sbh1 construct using XbaI and PstI restriction sites to replace Sbh1. *Hs*Get3/TRC40 was cloned into the MBP-Thr-*Ss*Get3 construct with XbaI and EcoRI (replacing the *Ss*Get3) to avoid an internal PstI restriction site.

Co-expression

BL21Gold(DE3) cells (Stratagene) were co-transformed with various Get3 constructs and TA proteins as follows: His-tagged TA proteins (pACYCDuet) co-transformed with MBP-Thr-Get3 (pMAL-C2) or untagged Get3 (pET33b) and MBP-tagged TA proteins (pMAL-C2) co-transformed with His-tagged or untagged Get3 (pACYCDuet). Co-transformants were selected with antibiotic resistant markers from both plasmids, inoculated into 2×YT medium cultures and grown at 37°C. Once the cultures reached an OD₆₀₀ of 0.6–0.8 they were induced with 0.3mM IPTG and allowed to express for 3–4h at 37°C. Cells were harvested by centrifugation at 6000g for 20m at 4°C.

Purification

Cells were resuspended in 10ml lysis buffer (50mM Tris pH 7.5, 300mM NaCl, 10mM βME and 0.1mM PMSF) per 1g cells using a Dounce homogenizer and lysed by 2–3 passes through

a ML-110L microfluidizer (Microfluidics). Cell debris was pelleted at 45000rpm in a Ti45 rotor for 30m at 4°C using an Optima L-80 XP ultracentrifuge (Beckman Coulter). The supernatant was applied to either Ni-NTA resin (Qiagen) or amylose resin (NEB) for 1-2h at 4°C. The resin was washed with 50mM Tris pH 7.5, 300mM NaCl, 10mM β ME (10mM imidazole when using Ni-NTA resin) and eluted with the same buffer containing 200mM imidazole or 10mM maltose for Ni-NTA and amylose resins, respectively. If thrombin was used for cleavage, 2U thrombin per ml of eluate was added and incubated for 12-16h at room temperature. When a second affinity purification step was used, the same protocol for the first step was followed but for the other affinity tag. Purified protein was concentration for injection on a Superdex 200 10/300 or 16/60 size-exclusion column (GE Healthcare) and run in 10mM Tris pH 7.5, 100mM NaCl and 10mM β ME. Fractions were analyzed by SDS-PAGE for the presence of complex, pooled and concentrated to 10-15 mg/ml for crystallization trials.

Crystallization and X-ray Diffraction

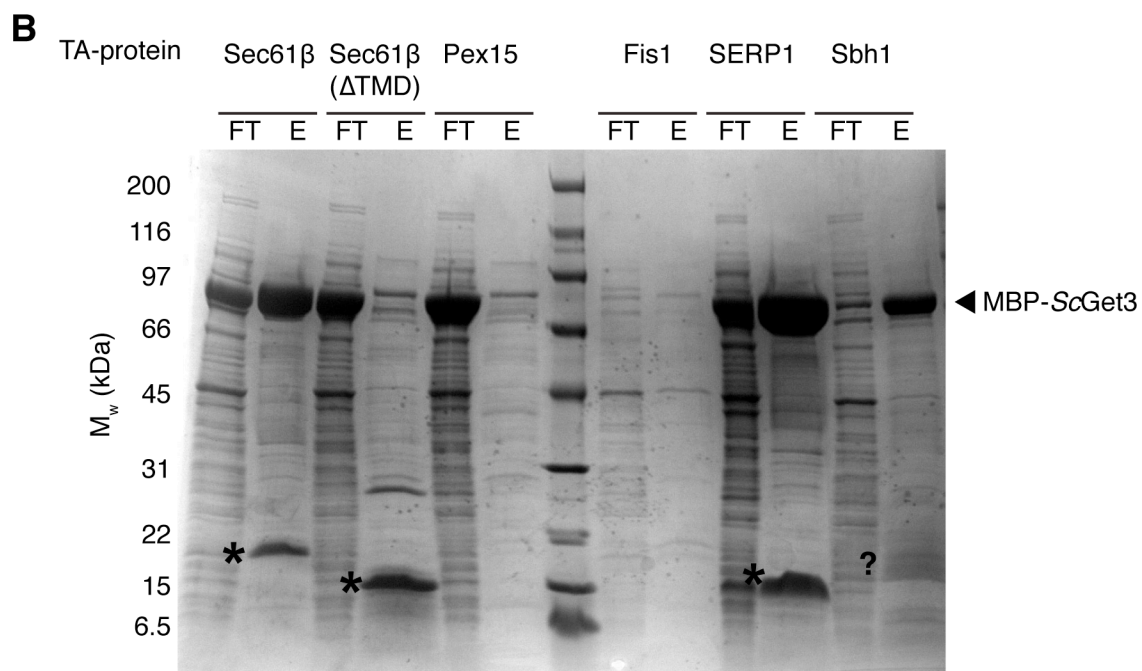
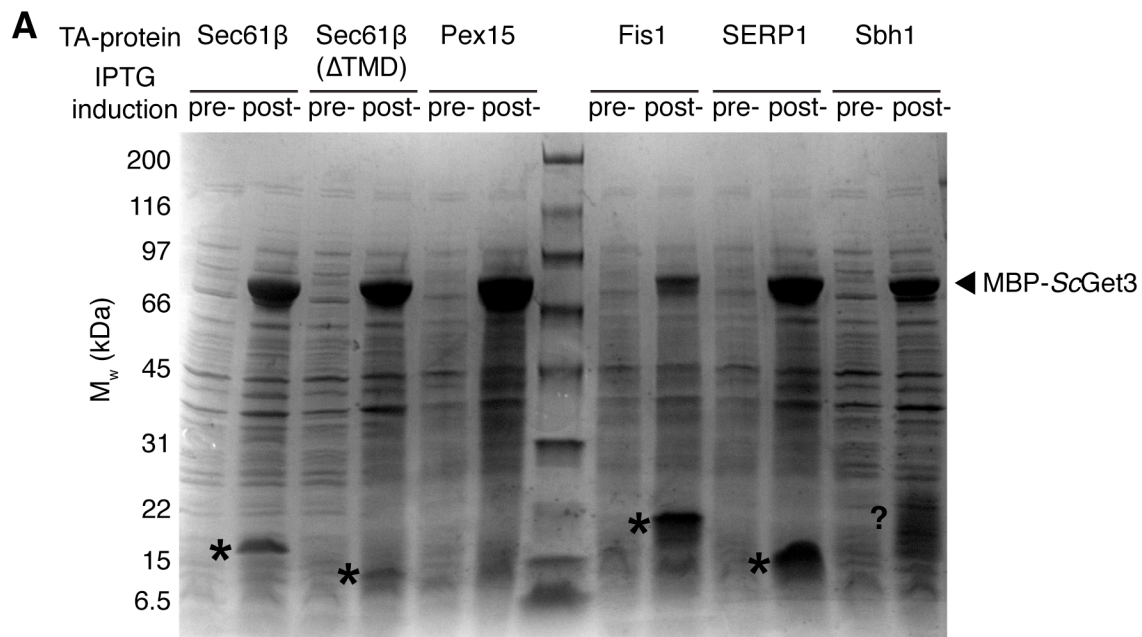
S₆Get3 complexed with Sbh1, Sbh2 or Ysy6 was dialyzed in 10mM Tris pH 7 and 10mM β ME for ADP·AlF_x reactions, and 2mM MgCl₂, 2mM ADP, 2mM AlCl₃ and 8mM NaF were added and incubated at room temperature for 1h. Complex used in AMPPNP reactions was dialyzed in 10mM HEPES pH 8 and 10mM β ME before addition of 2mM MgCl₂ and 2mM AMPPNP. Crystallization with the Crystal (Hampton), Index (Hampton), PEG/Ion (Hampton), MembFac (Hampton), PACT (Molecular Dimensions), JCSG+ (Qiagen) and Wizard (Emerald BioSystems) screens by sitting drop vapor diffusion were setup with equal volumes of protein and well solution (either 0.1 μ l or 0.2 μ l each) in MRC 2-well crystallization plates (Swissci) by a Mosquito liquid handling robot (TTP LabTech) at room temperature. Crystal hits were refined with a gradient of each component around the initial condition with

0.4, 1 or 2 μ l drops. Crystals from refined conditions (see Results section) were transferred to an artificial mother liquor containing either MgADP·AlF_x or MgAMPPNP and 20% glycerol or ethylene glycol (EG) for 5m, or they were soaked for 5m in 0, 5, 10 and then 20% glycerol or EG. Cryoprotected crystals were then flash frozen in liquid nitrogen and diffracted at 100K at the Stanford Synchrotron Radiation Lightsource, beam line 12-2, in the SLAC National Accelerator Laboratory. Diffraction data were collected using Blu-Ice (McPhillips et al., 2002) and analyzed with iMosflm (Battye et al., 2011).

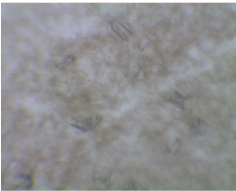
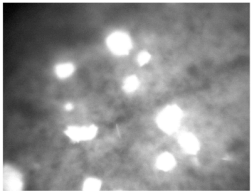
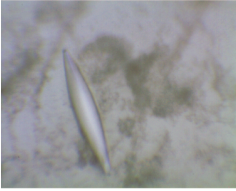
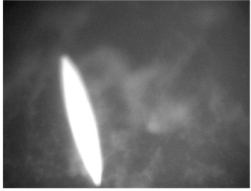
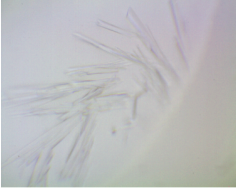
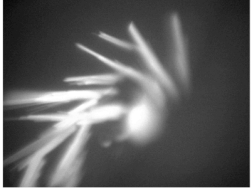

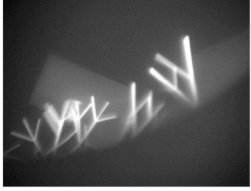



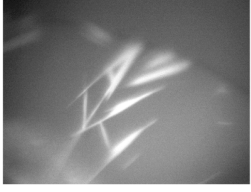


In vitro Reconstitution


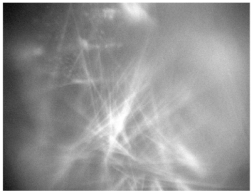
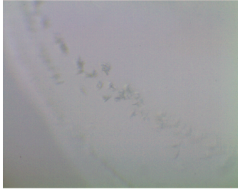
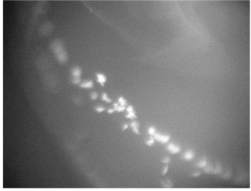
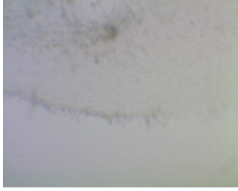
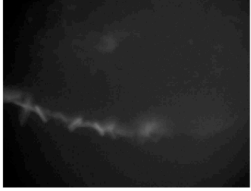
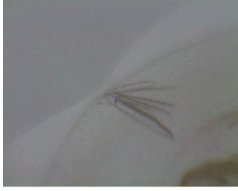
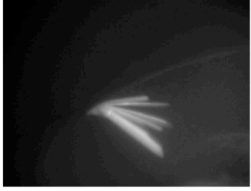

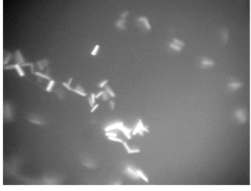



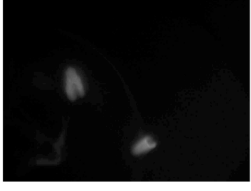
TA proteins for in vitro reconstitution were transformed into BL21Gold(DE3) cells, inoculated into 2×YT medium culture, grown to an at 37°C to an OD₆₀₀ of 0.6-0.8, induced with 0.3mM IPTG for for 3-4h at 37°C. Cells were harvested by centrifugation at 6000g for 20m at 4°C. Cell pellets were lysed with 50mM Tris-Cl, 5mM imidazole and either 8M urea or 6M guanidine HCl pH 7.5 at room temperature and spun down at 10000g for 30min. The supernatant was incubated on Ni-NTA resin for 1h, washed with the same buffer and eluted with buffer containing 200mM imidazole. Ultracentrifugation of samples was done at 60000g for 18m at 20C using a TLA 100.3 rotor in an Optima MAX-E ultracentrifuge (Beckman Coulter). Reconstitution by rapid dilution was performed by adding 150 μ l denatured TA protein drop-wise to 15ml of a stirred dilution buffer of 50mM Tris-Cl pH 7.5, 300mM NaCl, 4mM MgCl₂ and 10mM β ME and a 1:100 or 1:2 molar ratio of Get3 to added TA protein. Nucleotides were added to the buffer as 5mM ADP (ADP), 5mM ATP (ATP) or 5mM ADP, 5mM AlCl₃ and 8mM NaF (ADP·AlF_x). Complex was recovered by incubation with 100 μ l of Ni-NTA, washing twice with 500 μ l of dilution buffer with 10mM imidazole added and then eluting three times with 100 μ l of a dilution buffer with 250mM imidazole added.

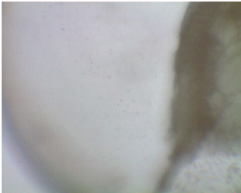
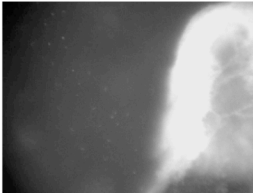

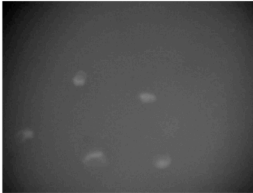
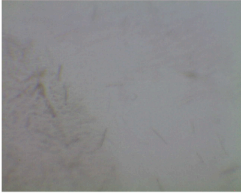

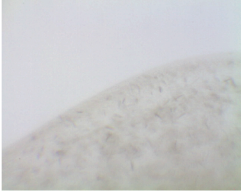
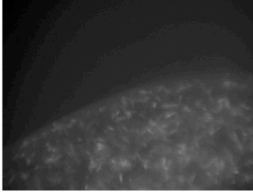
Supplementary Figures



Supplementary Figure 1. Recombinant co-expression and affinity purification of *S. Get3* complexed with TA proteins. (A) SDS-PAGE (coomassie-stained) of pre- and post-induction samples of an MBP fusion to *S. Get3* co-expressed with His-tagged TA proteins: Sec61 β , Sec61 β Δ TMD, Pex15, Fis1, SERP1 and Sbh1. The first lane is the pre-induction sample and the second is the post-induction sample, repeated for each TA protein. The position of Get3 is labelled and asterisks indicate bands corresponding to the expressed TA protein. Pex15, a peroxisomal protein, fails to express and the bands for Sbh1 are ambiguous as denoted by a question mark. Sec61 β , Sec61 β Δ TMD, Fis1, SERP1 express clearly in the post-induction samples. (B) SDS-PAGE (coomassie-stained) of affinity purification steps of Get3 complexed with TA-protein. Co-expressed complex was affinity purified for the His-tagged TA protein with Ni-NTA resin. Complex is retained for all expressed samples except the mitochondrial TA protein Fis1 which is insoluble and Sec61 β Δ TMD, both negative controls for specific TA protein binding to Get3.

Condition	visible	UV
ScGet3/Sbh1 (ADP•AlF _x) 0.1M HEPES pH 7.5 2% (v/v) PEG 400 2.0M Ammonium sulfate Crystal D3		
ScGet3/Sbh1 (ADP•AlF _x) 0.2M Sodium citrate tribasic dihydrate 20% (w/v) PEG 3350 Index H10 (1)		
ScGet3/Sbh1 (ADP•AlF _x) 0.2M Sodium citrate tribasic dihydrate 20% (w/v) PEG 3350 Index H10 (2)		
ScGet3/Sbh1 (ADP•AlF _x) 0.2M Ammonium tartrate dibasic 20% (w/v) PEG 3350 PEG/Ion D2		
ScGet3/Sbh1 (ADP•AlF _x) 0.2M Sodium citrate tribasic dihydrate 20% (w/v) PEG 3350 PEG/Ion D10		
ScGet3/Sbh1 (ADP•AlF _x) 0.2M Sodium/potassium tartrate 20% (w/v) PEG 3350 PACT E9		
ScGet3/Sbh1 (ADP•AlF _x) 0.2M Sodium citrate 20% (w/v) PEG 3350 PACT E11		

Condition	visible	UV
ScGet3/Sbh1 (ADP•AlF _x) 0.2M Sodium citrate 0.1M Bis-tris propane pH 6.5 20% (w/v) PEG 3350 PACT F11		
ScGet3/Sbh1 (ADP•AlF _x) 0.2M Sodium malonate 0.1M Bis-tris propane pH 6.5 20% (w/v) PEG 3350 PACT F12		
ScGet3/Sbh1 (ADP•AlF _x) 0.2M Ammonium citrate 20% (w/v) PEG 3350 Index Screen H10		
ScGet3/Sbh1 (ADP•AlF _x) 0.2M Ammonium sulfate 0.1M Sodium acetate pH 4.6 10% (w/v) PEG 4000 MembFac A3		
ScGet3/Sbh2 (ADP•AlF _x) 0.1M MMT buffer pH 7 25% (w/v) PEG 1500 PACT D4		
ScGet3/Sbh2 (ADP•AlF _x) 0.2M tri-Potassium citrate pH 8.3 20% (w/v) PEG 3350 JCSG+ B12		
ScGet3/Ysy6 (ADP•AlF _x) 0.2M Sodium citrate 0.1M Bis-tris propane pH 7.5 20% (w/v) PEG 3350 PACT E11		

Condition	visible	UV
ScGet3/Sbh1 (ADP•AlF _x) 0.2M Sodium fluoride 20% (w/v) PEG 3350 PACT E1		
ScGet3/Sbh1 (ADP•AlF _x) 15% (v/v) ethanol 0.1M Tris pH 7.0 Wizard D6		
ScGet3/Sbh1 (ADP•AlF _x) 0.1M Sodium citrate pH 5.6 20% (v/v) 2-propanol 20% (w/v) PEG 4000 Crystal D4		
ScGet3/Sbh1 (AMPPNP) 1.2M Sodium phosphate monobasic 0.8M Potassium phosphate dibasic 0.1M CAPS pH 10.5 0.2M Lithium sulfate Wizard C3		

Supplemental Figure 2. Crystals from initial screens of ScGet3 complexed with TA proteins. The condition lists the complex and nucleotide used followed by the composition of the condition and the screen and location it originated from. Images from visible and UV illumination of the crystals from the listed condition under a microscope is shown on the right.

Chapter 3

TAIL-ANCHORED PROTEIN TARGETING BY A GET3 TETRAMER

Introduction

Targeted delivery of membrane proteins is a highly regulated process. The ubiquitous co-translational signal recognition particle (SRP) pathway is responsible for the delivery of the majority of membrane proteins (Shan and Walter, 2005). Tail-anchored (TA) membrane proteins are exceptions. They are defined topologically as having a single transmembrane domain (TM) near their C-terminus (Borgese, 2003). Found in cytoplasmically associated membranes of all organisms, they account for 2–3% of open reading frames in humans and nearly 1% in yeast and prokaryotes (Beilharz et al., 2003; Kalbfleisch et al., 2007; Kriechbaumer et al., 2009; Pedrizzini, 2009; Borgese and Righi, 2010). The signal for TA protein membrane delivery, the C-terminal TM, is not accessible for targeting by the SRP pathway (Kutay et al., 1993); therefore, they must be delivered via a different route.

In eukaryotes, pathways for TA protein delivery to the ER have recently been elucidated (reviewed in Rabu et al., 2009; Simpson et al., 2010). The majority of TA proteins are targeted via the GET pathway (Guided Entry of TA proteins). This pathway, here described for yeast, progresses from a Sgt2/Get4/Get5 sorting complex (Bag6 complex in mammals (Mariappan et al., 2010)) that delivers an appropriate TA protein to the ATPase Get3 (TRC40 in mammals) (Battle et al., 2010; Chartron et al., 2010; Wang et al., 2010), which then targets the protein to the ER membrane via Get1/Get2 (Schuldiner et al., 2008). The central protein Get3 was the first component discovered that directly participates in TA-targeting (Stefanovic and Hegde, 2007). Get3 binds specifically to the TM of a TA-substrate and is essential for efficient delivery to the ER (Stefanovic and Hegde, 2007; Favaloro et al., 2008; Schuldiner et al., 2008).

Based on structural studies, a functional model for Get3 has been proposed where nucleotide state modulates an ‘open’ versus ‘closed’ homo-dimer (Bozkurt et al., 2009; Hu et al., 2009; Mateja et al., 2009; Suloway et al., 2009; Yamagata et al., 2010). As proposed, the ‘closed’ dimer uses a helical sub-domain (HSD) to form a hydrophobic groove for binding the TM of the TA protein. However, biochemical studies in mammalian extracts showed that TA proteins form complexes with Get3 compatible with a higher order complex (Stefanovic and Hegde, 2007; Favaloro et al., 2008). Furthermore, recombinant expression of Get3 with a TA-substrate yields a complex capable of TA membrane insertion *in vitro*. In this case, the complex contains a Get3 tetramer suggesting this is the oligomeric state of the targeting complex (Bozkurt et al., 2009; Favaloro et al., 2010). The difference between the dimer and tetramer models suggested by either structure or biochemistry remain to be reconciled.

A homologue of Get3 was recently identified in archaea that had previously been annotated as an ArsA, a structurally related bacterial ATPase involved in arsenate export (Borgese and Righi, 2010). Based on homology alone, distinguishing an ArsA from a Get3 is difficult; however, several key differences have been identified. The simplest is that Get3 is a homo-dimer while the ArsA monomer contains a tandem repeat, forming a pseudo-dimer. Get3 lacks the identified metal coordinating residues of ArsA (Boskovic et al., 1996; Stefanovic and Hegde, 2007) but contains a unique “Get3 motif” that is required for TA-binding (Mateja et al., 2009) (Figure S1). Additionally, Get3 homologues typically contain a pair of cysteines at their dimer interface that coordinate zinc and are essential for function (CxxC motif) (Metz et al., 2006; Stefanovic and Hegde, 2007).

The need for a specific TA-targeting protein, such as Get3, was thought to be unique to eukaryotes that contain membrane bound organelles and, thereby, multiple membranes for

insertion. The presumption has been that in prokaryotes there is no specialized machinery for delivery of TA proteins, as they would not require targeting to a specific membrane, indeed none have been found in eubacteria. The identification of a Get3 homologue in archaea brings this idea into question. Of the currently sequenced archaeal genomes roughly 50% contain a putative Get3. These can be classified into two groups based on the presence of the CxxC motif. They are found in methanogens, halophiles and thermophiles implying that organisms in these extreme environments have an additional level of complexity in membrane insertion.

Here we present the first structure of an archaeal Get3 from *Methanocaldococcus jannaschii* (*Mj*Get3). The structure is of a symmetrical homo-tetramer and features a hydrophobic chamber that we postulate sequesters the hydrophobic TA. We demonstrate that the archaeal Get3 is capable of TA binding, that the solution state of a fungal Get3/TA complex is similar to the tetrameric *Mj*Get3 and this fungal tetramer complex is capable of membrane insertion. This allows us to postulate a new model for TA-targeting by Get3.

Results

Structure of MjGet3

MjGet3 was purified and crystallized after recombinant expression in *E. coli*. We noted that the protein appeared to be a tetramer by size exclusion chromatography (SEC), further discussed below. The best crystals grew in the presence of ADP or the nucleotide analog ADP•AlF_x in two space groups. The two crystal forms were P2 diffracting to 3.2Å and P2₁ diffracting to 3.3Å grown in ADP and ADP•AlF_x respectively. Both structures were solved by molecular replacement, the P2₁ structure using a nucleotide-hydrolase domain (NHD) from *AjGet3* (3IBG) as the search model and the P2 structure using a NHD from the P2₁ structure. The remainders of both structures were built independently. The final refined structures had free R-factors of 29.6% for the P2 and 28.2% for the P2₁ forms. Data collection statistics are presented in Table S1. The two structures are very similar with an RMSD of 0.8Å over all Cα. There are four copies of *MjGet3* in the asymmetric units forming a homo-tetramer (Figure 1A & B, S2). The overall structure results in a dumbbell shaped particle approximately 150Å long. Unless noted, all figures will use the structure from the P2₁ crystal form.

Individual subunits of *MjGet3* are very similar to those from fungal Get3 structures (Figure 1C and S3). They closely align in the NHD with an RMSD of 0.8Å for residues 24-96, 149-171 and 234-333 of *MjGet3* to the transition state ‘closed’ *ScGet3* (2WOJ). There is clear density for ADP and a magnesium ion; however, at this resolution we could not clearly resolve the presence of an AlF_x in the P2₁ form (Figure 1D). There are a number of features of the NHD including switch I and II loops, which are responsible for transmitting changes in the nucleotide state to conformational changes related to function (Sprang, 1997; Leipe et al., 2002). For Get3, it has been suggested that they play a role in modulating between the open and closed state of Get3 as a dimer (Bozkurt et al., 2009; Mateja et al., 2009; Suloway et al.,

2009). Here, the switch I region is similar in conformation to structures of ‘closed’ fungal Get3 (Figure 1C and S3). Switch II is connected to helix 7. This helix has moved away from the NHD to form the tetramer cage pulling the switch II loop into a conformation not compatible with hydrolysis (Figure 1D and S4A & B). The organization of this region is closer to the conformation of the structure of the ‘closed’ fungal Get3 bound to ADP (3IQX) (Figure S3).

As seen before, two of the monomers come together to form a dimer stabilized by a cysteine-coordinated zinc (Figure 1A & B). There are two of these dimers in each tetramer. The structures are very similar and were partially constrained by non-crystallographic symmetry during refinement. The orientations of the monomers in the dimer are most consistent with the transition state ‘closed’ form of *Ss*Get3 (Figure 1E & F and S3).

The most dramatic difference between tetrameric *Mj*Get3 and the dimeric ‘closed’ *Ss*Get3 are the helices that surround the putative TA binding groove (Figure 1E & F). Helix 6, which lies at the base of the groove, matches the conformation of the transition state ‘closed’ *Ss*Get3 (2WOJ), tilted relative to other fungal structures. Helix 6 is shorter in the *Mj*Get3 tetramer than in the *Ss*Get3 dimer and has moved in the direction of the dimer interface. This shortens the proposed hydrophobic binding groove by $\sim 10\text{\AA}$ (Figure S3) (Mateja et al., 2009). The end of helix 6 becomes more exposed relative to the ‘open’ form (2WOO) (Figure S3) consistent with hydrogen/deuterium exchange experiments that show this helix exchanges hydrogen more rapidly after TA binding (Bozkurt et al., 2009). The five helices flanking the groove (helices 4,5 and 7,8,9) are in a similar orientation to that seen in the ‘closed’ fungal structures (Figure 1E & F). They are extended and more ordered, with only eight residues connecting helix 8 to helix 9 missing backbone density.

Instead of a hydrophobic binding groove, the *Mj*Get3 tetramer uses the flanking helices to form the walls of a cage generating a hydrophobic central chamber (Figure 1A & B). Here, and in closed fungal Get3 structures, the loop formed by helices 4 and 5 (4/5 loop) is on the opposite side of helix 6 from the loop formed by helices 7 through 9 (7/8/9 loop) (Figure 1C). In *Mj*Get3 the 4/5 loop tilts away from the binding groove. In doing so it forms a three-helix bundle to helix 8 extended from a subunit across the tetramer stabilized by hydrophobic interactions (Figure 1A & B). Calculated by PISA (Krissinel and Henrick, 2007), the tetramer buries an additional $\sim 6000 \text{ \AA}^2$ relative to the dimers ($\sim 11,000 \text{ \AA}^2$) with the three-helix bundle providing most of the contacts.

The chamber formed by the tetramer is predominantly hydrophobic (Figure 2A). The size of the internal cavity is $\sim 30 \text{ \AA}$ across the middle to $\sim 40 \text{ \AA}$ down the long axis (Figure S5). This is compatible with the dimensions of a single TM with volume remaining for additional helices. The chamber is lined by helices 5, 7, 8 & 9. Unlike in the dimer model, helix 4 is on the periphery as part of the three-helix bundle (Figure 2B). The two helix 8s from the dimer extends into the groove of the opposing dimer (Figure 2B). The result is that they cover the floor and block the ends of the groove. This configuration prevents direct contact from helix 6 to the chamber. The cage has openings at the site where the disordered loop between helix 8 and 9 is missing (Figure 2C), which could provide an access point between the internal chamber and the cytoplasm. This opening, when viewed based on electrostatic potential, has an overall positive charge. The general charge is conserved in eukaryotes and is consistent with a discrimination point preventing binding of mitochondrial TA proteins (Mateja et al., 2009).

Extensive mutagenesis has been performed on *Sa*Get3 by our lab and others (Figure 2D and S6) (Mateja et al., 2009; Suloway et al., 2009). Mateja et al. focused on the putative TA-binding groove and identified three classes of mutations: those that affect TA-binding (blue), nucleotide hydrolysis (red) or both (purple). The location of residues that affect nucleotide hydrolysis predominantly cluster to the base of the groove on helix 6; however, in the dimer model these residues did not affect TA-binding despite being components of the putative groove. In the tetramer structure, the reason for this becomes clear as these residues do not directly contact the hydrophobic chamber and, therefore, would not be predicted to affect binding (Figure S6B & C). They do lie directly below the chamber and would be expected to communicate the state of the chamber to the NHD. The mutants that affect TA-binding all coat the interior of the hydrophobic chamber (Figure 2D & S6C).

Archaeal Get3 Binds Tail-anchored Proteins

The high sequence and structural homology between fungal and archaeal Get3 suggests that the archaeal protein is capable of binding TA proteins. We established a method for purifying a Get3/TA protein complex heterologously in *E. coli* using a two-step purification procedure where both Get3 and the TA protein contain affinity tags (Figure 3A) (similar to (Bozkurt et al., 2009; Favaloro et al., 2010; Yamagata et al., 2010)). Using this method we could reliably purify *Sa*Get3 bound to a variety of TA-substrates. We analyzed the complex using inductively coupled plasma mass spectrometry (ICP-MS) to identify bound ions (Yamagata et al., 2010). We found an approximate 2:1 ratio of Get3 to zinc but only trace amounts of magnesium suggesting that there are no appreciable amounts of bound nucleotide, consistent with what has been reported before. An example of this complex is the purification of *Sa*Get3 bound to Sbh1, the yeast homologue of mammalian Sec61 β and a demonstrated GET pathway substrate (Figure 3B, lane 1) (Schuldiner et al., 2008). Interestingly, when we co-expressed

*Mj*Get3 with Sbh1 we were able to purify a stable chimeric complex confirming that the archaeal homologue is capable of binding a TA protein (Figure 3B, lane 2). As a control, co-expression of *Mj*Get3 with a TM deletion of Sbh1 was unable to form a complex by this method demonstrating that *Mj*Get3 specifically bound the TA (Figure S7A & B).

Next, we tested to see if the archaeal Get3 was capable of binding a native substrate. We first co-expressed *Mj*Get3 with *Mj*Sec β , the archaeal homologue of Sec61 β whose structure has been solved and is predicted to be a TA protein (Van Den Berg et al., 2004; Borgese and Righi, 2010). Using our two-step purification, we were able to obtain a stable complex of the two proteins (Figure 3B, lane 3). Neither component was recovered when expressed alone and purified by the same two-step procedure.

A number of additional TA proteins have been identified bioinformatically from *Methanococcus maripaludis*, a related species to *M. jannaschii* (Borgese and Righi, 2010). Some of these have homologs in *M. jannaschii* including SecE, another Sec channel component and MtrA and MtrB, TA-subunits of tetrahydromethanopterin S-methyltransferase. All three proteins could be purified as a complex bound to *Mj*Get3 (Figure 3B, lane 4-6). We wanted to test another archaeal homologue and chose to work with the Get3 from *Thermococcus kodakaraensis* (*Tk*Get3). This Get3 falls into the class of archaeal homologues missing the pair of zinc-coordinating cysteines. Using the two-step method, *Tk*Get3 is capable of forming stable complexes with the same set of TA proteins tested for *Mj*Get3 (Figure 3B, lanes 7-11).

The binding chamber only sequesters the TM

Get3 specifically recognizes and binds the hydrophobic TA (Stefanovic and Hegde, 2007). The nature of the tetramer's hydrophobic chamber suggests it is capable of sequestering a TA; however, it would be unable to fit a typical soluble domain. We would postulate that while the

TA is in the chamber the soluble domain extends out through the gap in the cage. The consequence would be that Get3 would only protect a minimal amount of the TA protein from solvent. Our double affinity tagged system provides a means to test this. Here, we took the N-terminal fusion of MBP to Sbh1 and introduced a thrombin protease site between the two proteins, which we then co-expressed with *Sa*Get3 and purified (Figure 3A & C). For the full-length MBP-Sbh1 fusion, the addition of thrombin resulted in a gel shift of MBP from running as a fusion to MBP alone (Figure 3D, lane 1 versus lane 2). TM topology predicting software, TMHMM 2.0 (Krogh et al., 2001), predicts the TM to extend from residue 55 to 74 (Figure 3C). We generated constructs with progressively shorter N-terminal soluble domains of Sbh1 and tested them in the assay. A complex of *Sa*Get3 with an MBP fusion of Sbh1 truncated to eight amino acids N-terminal of the TM could be cleaved (Figure 3D, lane 5 versus 6); however, Sbh1 truncated to three amino acids N-terminal to the TM could not (Figure 3D, lane 7 versus 8). Therefore, Get3 sequesters only a few amino acids in addition to the hydrophobic TM. This result could also be consistent with the dimer model where the groove covers only the TA.

Get3 can form a tetramer in solution

The current model for the function of fungal Get3 is based on Get3 always maintaining a dimeric state. *Sa*Get3 purified after expression in *E. coli* is predominantly a dimer by SEC; however, a small pool always purified as a tetramer (Figure 4A & B, solid blue trace). This tetramer pool was stable enough to be re-run over the column (Figure 4B, dashed line). We noted that the tetramer fraction would degrade over time to dimer, while we would never see conversion of the dimer to tetramer. We suspected, as noted below, that the tetramer fraction is stabilized by interactions with hydrophobic peptides. In this case, tetrameric *Sa*Get3 might be bound to hydrophobic *E. coli* peptides. We searched for evidence of these by mass

spectrometry but were unable to find any *E. coli* peptides. This may not be surprising, as the expected hydrophobic peptides are typically hard to identify by liquid chromatography tandem mass spectrometry (Wu and Yates, 2003) and may be in low abundance.

In contrast to *Sa*Get3, the solution state of *Mj*Get3 is primarily a tetramer by SEC (Figure 4A, green trace), which is consistent with the crystal structure. We were interested to see if this was a general feature of archaeal Get3 homologues. Get3 from the mesophile *M. maripaludis* (*Mm*Get3) was predominantly tetrameric in solution (Figure 4A, red trace). *Tk*Get3, which lacks the coordinating cysteines, is similar to *Sa*Get3 in that the protein elutes as both a tetramer and a dimer (Figure 4A, cyan trace).

The interactions forming the tetramer in the *Mj*Get3 crystal structure are primarily hydrophobic packing between helix 8 on one subunit and helices 4 and 5 from the opposing subunit (Figure 4C). This suggests that the tetramer should be sensitive to detergent. To test this, we dialyzed the protein against a 1% (w/v) solution of the small micelle detergent N-octyl- β -D-glucopyranoside (β -OG) just above the critical micelle concentration. After overnight dialysis the majority of the *Mj*Get3 shifted to a peak consistent with a dimer (Figure 4D, compare green trace to black trace). When we then dialyzed out the detergent, the protein returned to the tetramer (Figure 4D, black dashed trace) confirming that in these conditions this was the stable state. When we diluted out the detergent in the presence of ATP we saw no effect on the oligomeric state, although we did note an increase in absorption consistent with nucleotide binding (260/280 ratio went from 0.81 without nucleotide to 1.02 in the presence of ATP). This was intriguing, as adding ATP to purified tetrameric *Mj*Get3 would cause the protein to precipitate. For further confirmation of the stability and hydrophobic nature of the

interaction, we performed SEC of MjGet3 in the presence of high salt (1M NaCl) or denaturant (1M urea). Neither affected the oligomeric state (Figure S8).

Helix 8 had been suggested to play an important role in TA-binding, possibly as a cover to the groove proposed in the dimer model (Mateja et al., 2009). Deletion of this helix resulted in a loss of TA-binding (Yamagata et al., 2010). In the tetramer model, this helix stabilizes the interface; therefore, it is critical to formation of the TA-binding hydrophobic chamber. To test this directly, we introduced mutations in helix 8 of *MjGet3* near the interface of the three-helix bundle (Figure 4C). Two of these, F192D and M196D, directly disrupt the hydrophobic interface and both of them shift predominantly to a dimer peak by SEC (Figure 4E, red and purple trace). A third mutation, M193D, was not at the interface (Figure 4C) and this mutant strongly favored the tetramer (Figure 4E, orange trace). Partial deletion of helix 8 resulted in a loss of the tetramer peak as well (Figure 4E, yellow trace).

Fungal Get3 tetramer is capable of TA membrane insertion in vitro

A biological role for the tetramer complex is difficult to prove directly. We decided to address this by demonstrating that our purified Get3/TA tetramer complex was capable of insertion into purified yeast microsomes. It had previously been shown that both a fungal Get3/TA tetramer complex (Bozkurt et al., 2009) and a mammalian Get3/TA tetramer complex (Favaloro et al., 2010) are capable of insertion into mammalian microsomes. In both cases, membrane integration was verified by glycosylation of the C-terminus of the TA protein, which can only happen if the C-terminus has entered into the lumen of the microsomes. We chose to perform a similar assay using all fungal components.

We generated an MBP-tagged Sbh1 with a glycosylation site from opsin at its C-terminus (MBP-Sbh1-op). This purified as a stable tetrameric complex with *Ss*Get3. Using this complex

we observed successful insertion of MBP-Sbh1-op into *S. cerevisiae* $\Delta get3$ microsomes by glycosylation of the C-terminal opsin tag, confirmed by subsequent deglycosylation with the endoglycosidase EndoH after disruption of the microsomes (Figure 5A, Lanes 3 & 4). As previously seen (Favaloro et al., 2010), no insertion was observed when microsomes were pre-treated with trypsin (Figure 5A, Lanes 1). Furthermore, insertion is sensitive to the binding of nucleotide, here disrupted by the addition of EDTA to the reaction (Figure 5A, Lane 2). MBP-Sbh1-op could be purified without Get3 and alone failed to insert (Figure 5A, Lane 5). This shows that the *Sc*Get3/TA tetramer complex is on a functional insertion pathway. We were interested to see if we could get transfer from an archaeal complex into our yeast microsomes. We purified a stable *Mj*Get3/MBP-Sbh1-op tetramer complex. This complex was unable to insert the yeast substrate into *S. cerevisiae* $\Delta get3$ microsomes (Figure S9A).

The inability of the archaeal *Mj*Get3/yeast TA tetramer complex to insert into yeast microsomes implies that there are significant differences in the putative archaeal pathway at the membrane. Indeed, there are no obvious homologues of either Get1 or Get2 in any archaea. To further explore the species dependence of our insertion results, we decided to use a reconstituted in vitro translation system that has previously been used to demonstrate insertion by eukaryotic Get3 variants (Stefanovic and Hegde, 2007). Using lysates from the *S. cerevisiae* $\Delta get3$ strain, we were able to reconstitute insertion into purified microsomes using a model substrate (N-terminal truncation of MBP-Sbh1-op) dependent on the presence of *Sc*Get3 (Figure S9B). Based on this, we decided to see if any of the archaeal Get3 homologues could facilitate insertion. Similar to the purified complex assay, *Mj*Get3 was unable to facilitate insertion, as was *Mm*Get3 or *Tk*Get3 (Figure S9B & C). This suggests that, under these conditions, archaeal Get3 homologues are unable to support insertion into fungal microsomes.

Solution characterization of Get3 and Get3-TA protein complexes

Complexes between fungal or mammalian Get3 bound to TA-substrates are consistent with a tetramer of Get3 (Bozkurt et al., 2009; Favaloro et al., 2010). We decided to assay the oligomeric state of various complexes using multi-angle laser light scattering (MALLS), which provides a relative molecular weight of the particle. Using this method, both the purified *Sa*Get3 dimer and the tetrameric *Mj*Get3 were consistent with calculated molecular weights, 80.8 kD and 153 kD respectively (Figure 5B and Table 1). Using purified *Sa*Get3/TA complexes the experimental molecular weights were consistent with a tetramer of *Sa*Get3; however, the stoichiometry of the TA proteins was inconclusive (Figure 5B and Table 1). This suggests that more than one TA protein is bound per Get3 tetramer possibly because the complex is formed in the absence of other GET partners. The size of the chamber should be able to accommodate multiple TA proteins (Figure S5).

The data are consistent with the *Mj*Get3 tetramer having a similar conformation to the Get3/TA complex. We expected that the elongated dumbbell structure and the stability of the purified complexes could be exploited in an analysis using biological small-angle X-ray scattering (bioSAXS). The benefit of this technique is that it can provide measures of dimensions in solution along with allowing for the calculation of low-resolution molecular envelopes (Putnam et al., 2007). bioSAXS curves of *Mj*Get3 and *Sa*Get3/TA complexes show similar dimensions such as the radii of gyration (R_g) and maximum dimension (D_{max}) (Table 1). The unbiased overall shapes of the pair-distribution functions, $P(r)$, are also similar with a primary peak followed by a shoulder peak (Figure 6A), strongly suggesting a multi-domain protein consistent with the crystal structure of *Mj*Get3. This further indicates that the fungal and archaeal tetramers have similar overall architectures.

In addition to overall dimensions, bioSAXS can use the experimentally measured distance distributions (Figure 6A) in a variety of refinement procedures to obtain ab initio molecular envelopes (Putnam et al., 2007). We used DAMMIN (Svergun, 1999) to calculate molecular envelopes of each of our complexes. Knowing that the Get3 tetramer has internal symmetry, we imposed P22 symmetry on our model. This did not affect the overall dimensions of any of the complexes; however, it was necessary to establish a consistent envelope. We calculated a molecular envelope for *Mj*Get3 (Figure 6B). The ab initio fit to the data returned a dumbbell shaped envelope consistent with the crystal structure. Viewed down the two-fold along the long axis of the tetramer, we measured a crossing angle of the widest point in each of the dimers. In the crystal structure this crossing angle is $\sim 30^\circ$. Performing a similar measure for the molecular envelope results in a crossing angle of $\sim 40^\circ$. This suggests that the crystal structure stabilizes a slightly twisted form of the *Mj*Get3 compared to the solution state. We generated a molecular envelope for both the *Sc*Get3/Ysy6 complex and a truncated *Sc*Get3/Sbh1₄₇₋₈₂ complex (Figure 6C & D). These resulted in very similar dumbbell shaped structures confirming that the *Mj*Get3 tetramer is a good model for the fungal Get3/TA tetramer complex.

Figures

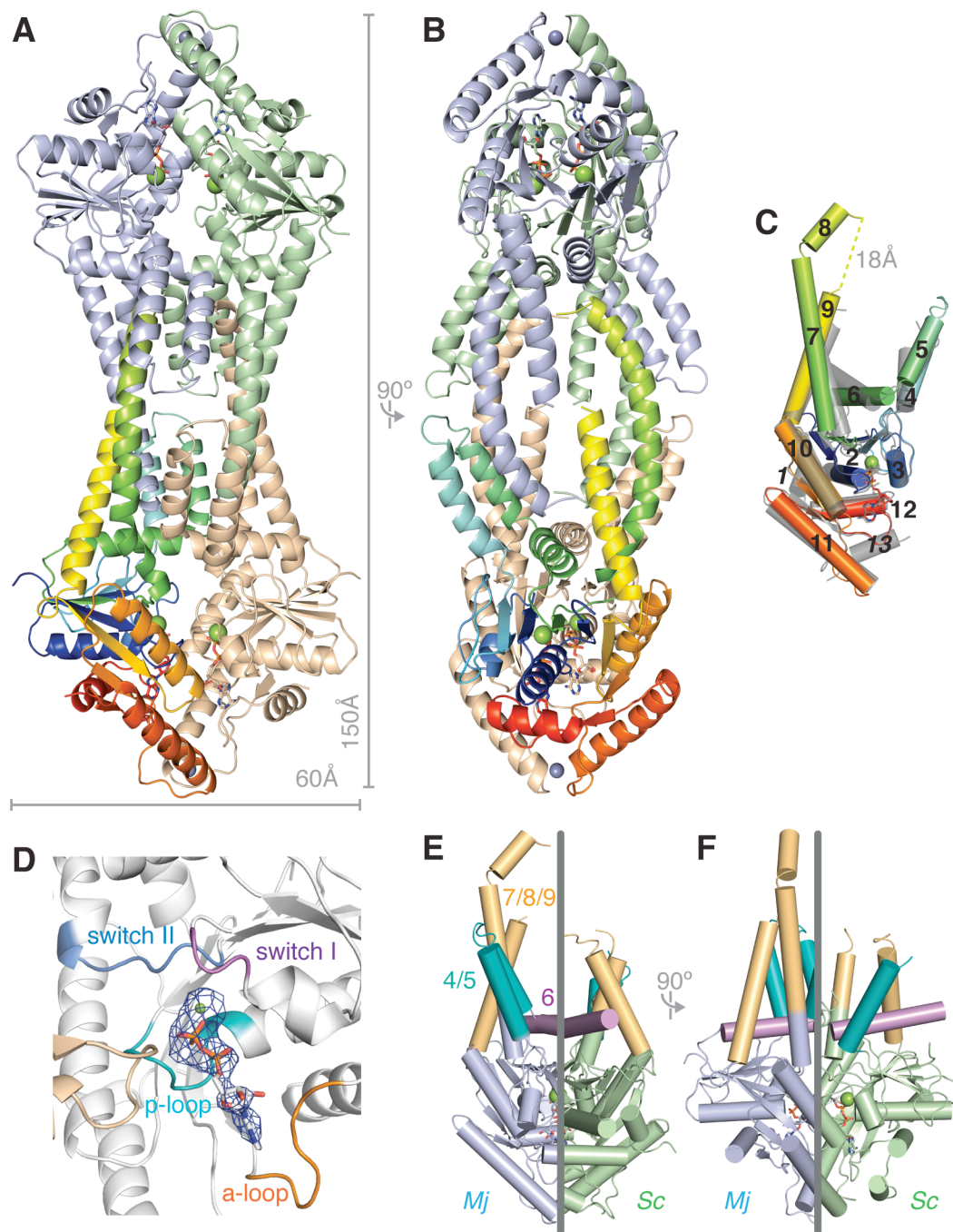


Figure 1. The structure of *Mj*Get3. (A) The overall structure of *Mj*Get3 in cartoon representation with each subunit colored differently and one subunit color ramped blue to red from N- to C-terminus with approximate measurements on the side. Nucleotides are represented as sticks and ions as spheres. (B) 90° rotation relative to A. (C) A monomer of *Mj*Get3 color ramped as in A overlaid with *Sc*Get3 (2WOJ-A) shown in gray. Helix 1 is not resolved in *Mj*Get3 and helix 13 is not obviously present. Dashed line connects helix 8 to helix 9. (D) A view of the nucleotide binding pocket highlighting hydrolase features: P-loop (cyan), switch I (purple), switch II (blue), and A-loop (orange). The opposing subunit in tan. ADP is in sticks colored by atoms. Mg²⁺ as green sphere. 2F₀-F_C density for the nucleotide is shown as a blue mesh contoured at 2.5σ. (E) A split-view comparison of the dimers of *Mj*Get3 (lavender) and *Sc*Get3 (2WOJ—green). The dimers each have 2-fold symmetry in the views shown. For clarity, only half of each structure from the overlay is shown to give a direct comparison. *Mj*Get3 on the left and *Sc*Get3 on the right. Colored helices are 4/5 (teal), 6 (purple) and 7/8/9 (light orange). (F) A 90° rotation relative to E.

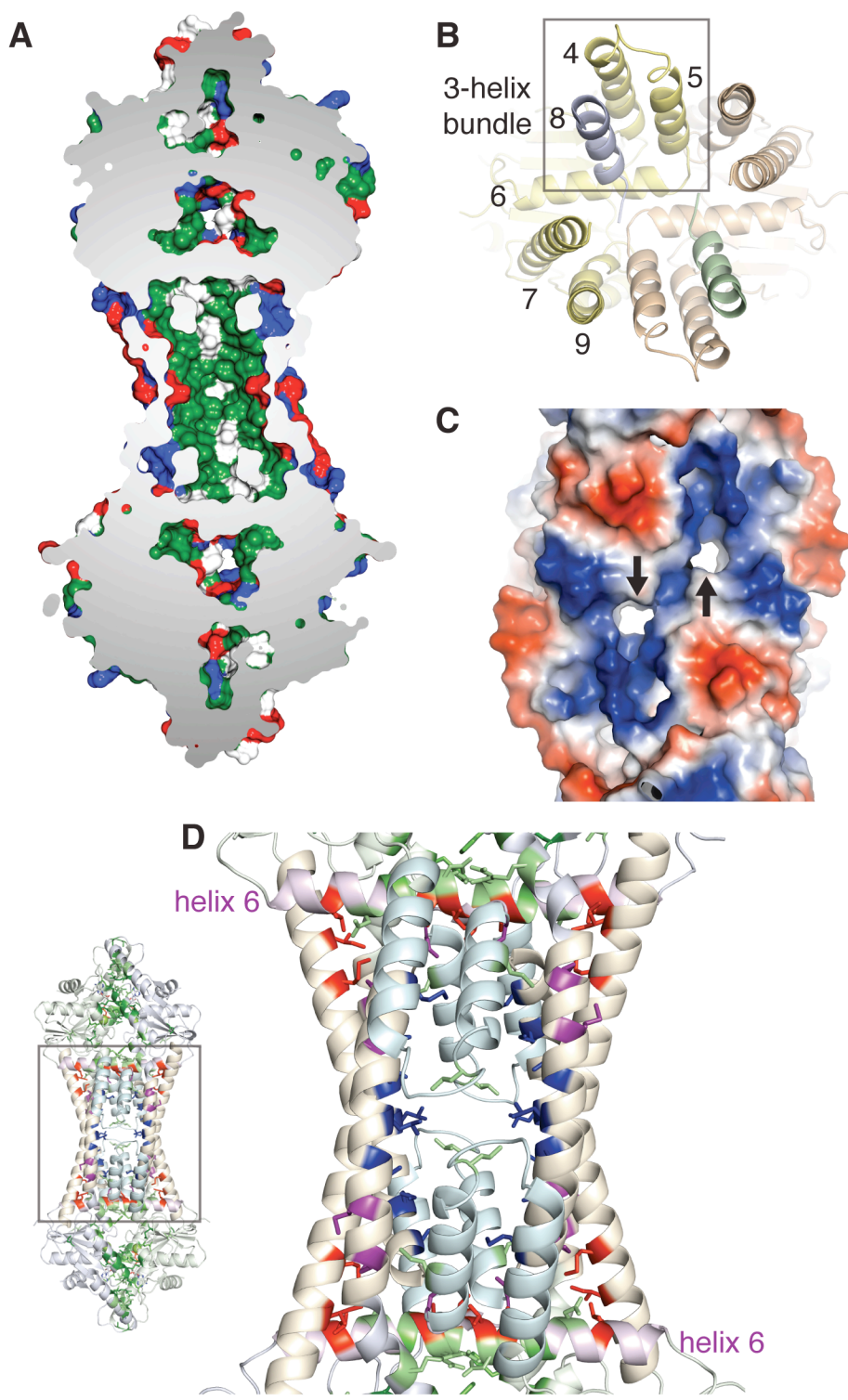


Figure 2. Central cavity. (A) Surface representation of *Mj*Get3 cut through the middle. Residues colored based on type: positive charge (blue), negative charge (red) and hydrophobic (green). (B) Similar to Figure 1A rotated 90° forward showing the cage walls with the foreground removed. (C) External view of the cavity in surface representation colored based on electrostatic potential (negative—red to positive—blue). Holes indicated by arrow. (D) External view of the central cavity highlighting mutants from previous studies, inset shows zoom in region colored similar to Figure 1E. Mutants resulting in a negative growth phenotype by (Suloway et al., 2009) are shown in green. Mutants from (Mateja et al., 2009) are colored according to loss of nucleotide hydrolysis (red), TA-binding (blue) or both (purple).

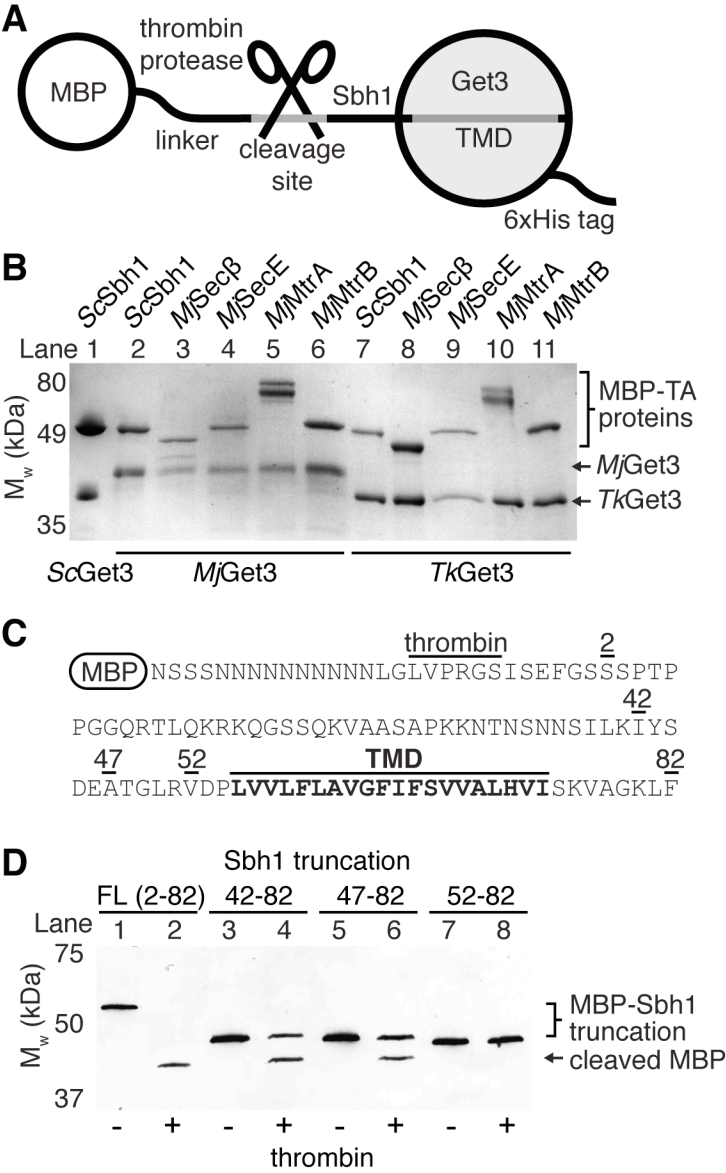


Figure 3. TA protein binding by Get3. (A) Diagram of recombinantly expressed complex showing the two affinity tags used for purification (MBP, maltose binding protein). Position of thrombin cleavage site indicated. (B) Coomassie-stained SDS-PAGE of Get3/TA protein complexes from various species purified by recombinant co-expression. (C) Sequence of Sbh1 fusion to MBP with residues from truncations indicated. The TMD is in bold. (D) SDS-PAGE and Western blot of the *S. Get3*/Sbh1 truncation complexes pre- and post-thrombin cleavage with α MBP antibody against the MBP-Sbh1 fusion. Accessibility of protease site results in a shift of the MBP-Sbh1 fusion to a lower MBP band. Residues numbers of Sbh1 are indicated.

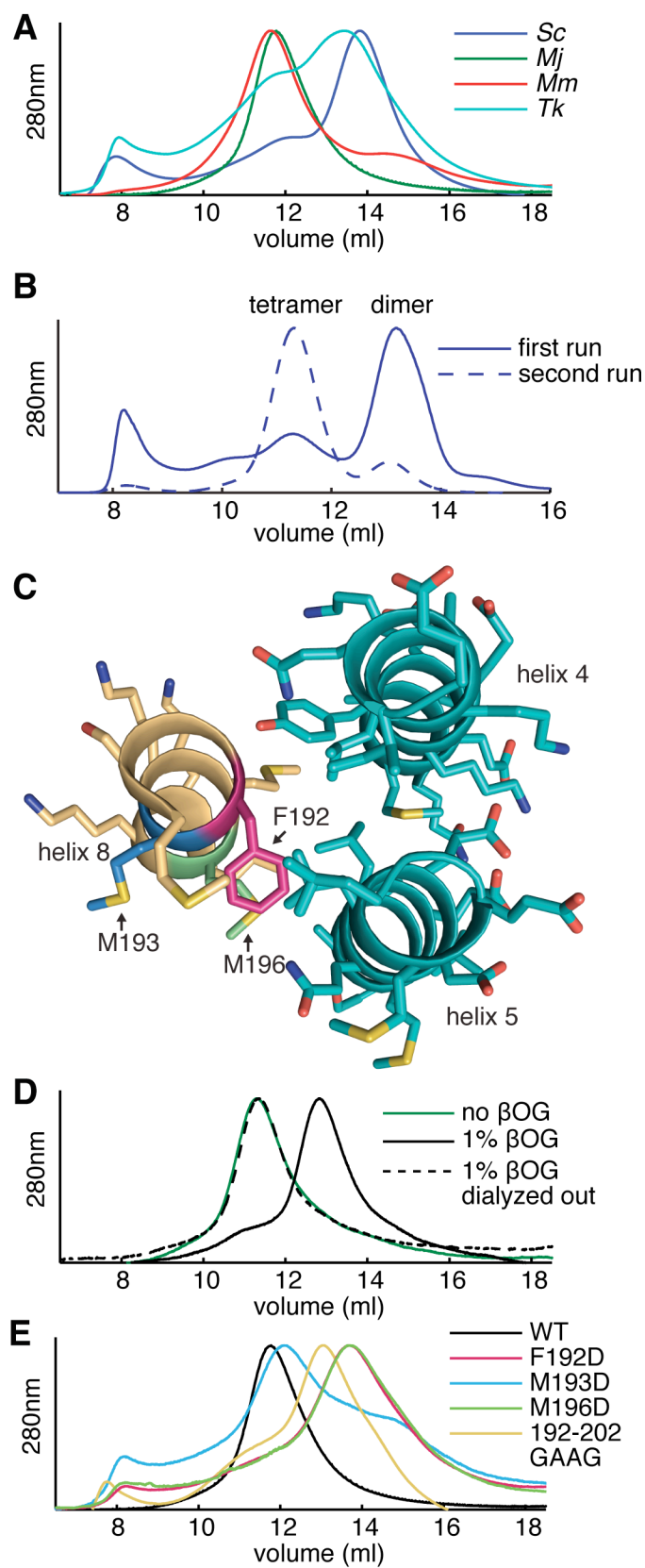


Figure 4. Oligomeric state of Get3. (A) SEC of Get3 from *S. cerevisiae* (blue) and three archaeal species: *M. jannaschii* (green), *M. maripaludis* (red) and *T. kodakaraensis* (cyan). Tetramers and dimers eluted around 12ml and 14ml respectively. (B) SEC run on different column of *Sc*Get3 after affinity purification (solid line). The tetramer peak was pooled and re-run (dashed line). (C) The three-helix bundle that stabilizes Get3 tetramers colored as in Figure 2B with mutated residues highlighted. (D) SEC of *Mj*Get3 in the absence (green) and presence (black) of 1% β -OG and a sample where detergent was dialyzed out (dashed). Detergent shifts the peak to a volume corresponding to a dimer (13ml). After dialysis the peak shifts back to a tetramer (11ml). (E) SEC of *Mj*Get3 mutants. The 192–202 GAAG trace (yellow) corresponds to a deletion of helix 8.

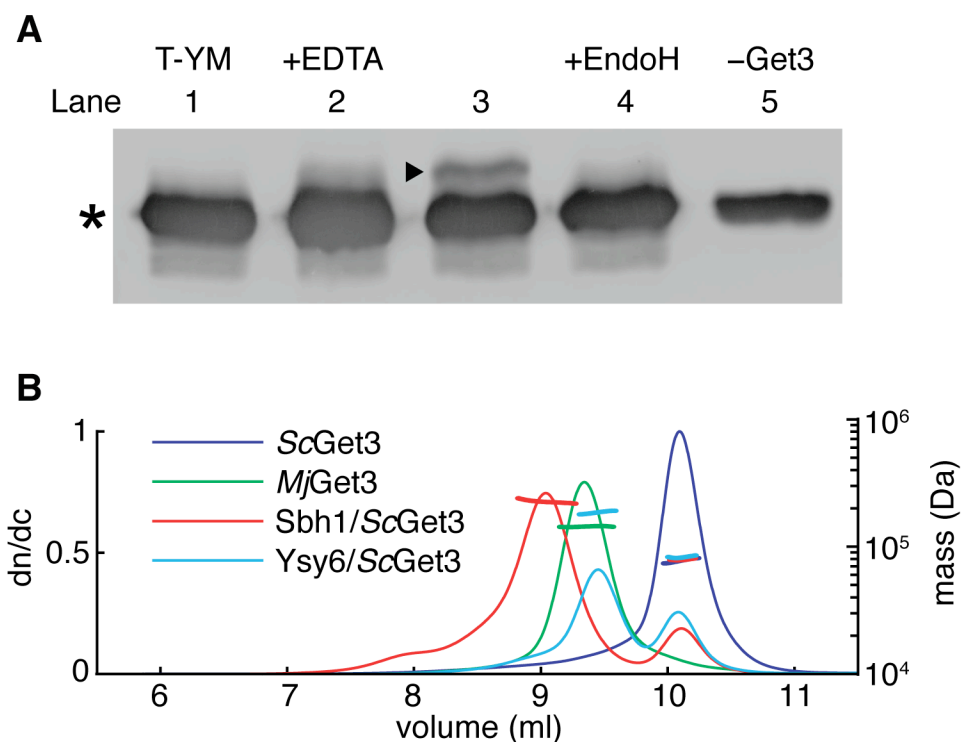


Figure 5. Functional studies of Get3/TA protein tetramer complex. (A) In vitro membrane integration of an MBP fusion to Sbh1 with a C-terminal opsin tag into microsomes purified from *S. cerevisiae* $\Delta get3$ by a purified ScGet3/MBP-Sbh1-op complex. A Western blot against MBP of in vitro translocation assays into trypsinized yeast microsomes (T-YM – Lane 1), in the presence of EDTA (Lane 2), standard in vitro translocation conditions before (Lane 3) and after EndoH treatment (Lane 4) and MBP-Sbh1 purified without Get3 (Lane 5). Star indicates MBP-Sbh1-op and the arrow points to band shifted by glycosylation of MBP-Sbh1-op after membrane integration. (B) Molecular weights of Get3 and Get3/TA protein complexes measured by SEC coupled to MALLS. Traces of differential index of refraction (dn/dc) and calculated molecular weights are shown.

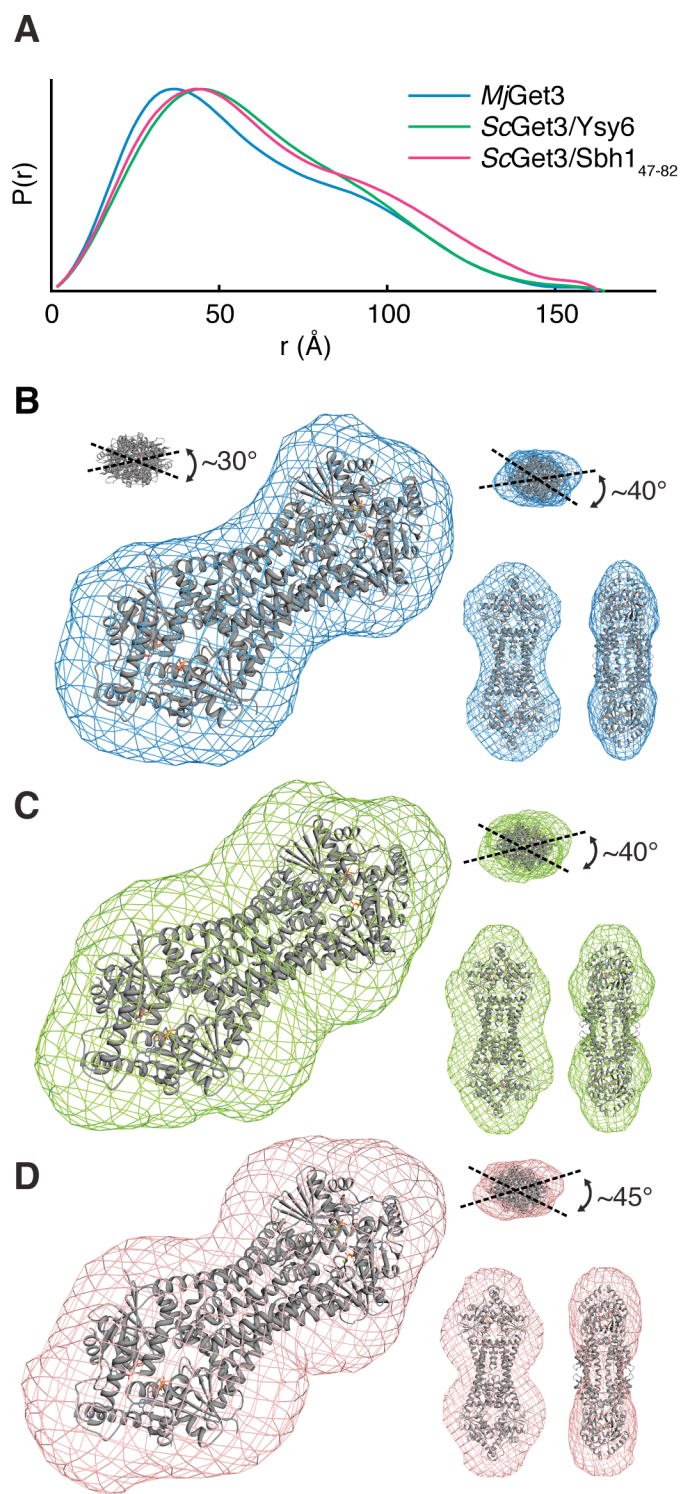


Figure 6. Size and shape of Get3/TA protein complexes. (A) Pair-distribution functions from bioSAXS of *Mj*Get3 (blue), *Ss*Get3/Ysy6 and *Ss*Get3/Sbh1₄₇₋₈₂. (B) Calculated envelope of *Mj*Get3 from bioSAXS data (blue mesh) with *Mj*Get3 coordinates fit to the overall envelope. Insets are three views 90° rotated around the vertical and horizontal axis. Lines are drawn through the widest point of each lobe and a crossing angle is calculated along the long axis. Top left inset shows the crossing angle for the structure alone. (C) Similar to B for the envelope of *Ss*Get3/Ysy6 complex (green mesh). (D) Same as C for the envelope of the *Ss*Get3/Sbh1₄₇₋₈₂ complex (salmon mesh).

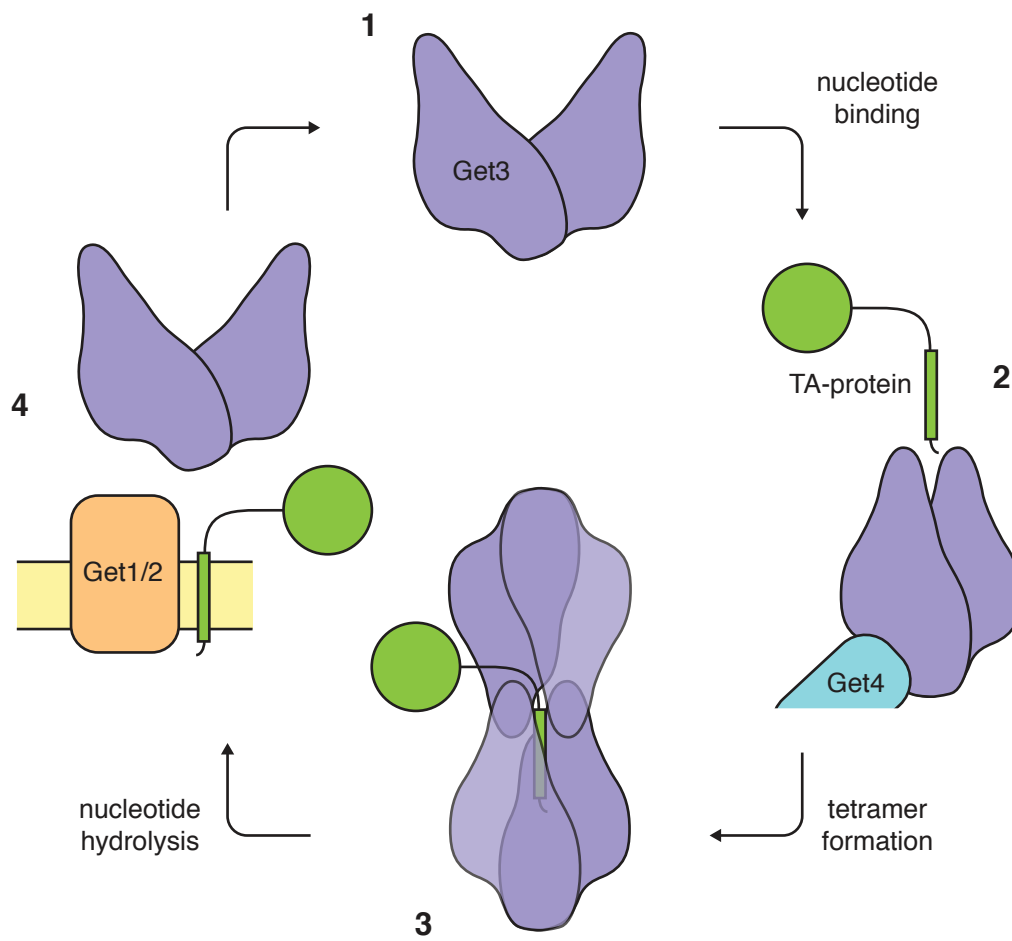


Figure 7. Model for Get3 TA protein targeting.

MALLS	<i>S</i> ϵ Get3	<i>M</i> γ Get3	<i>S</i> ϵ Get3/ Ysy6	<i>S</i> ϵ Get3/ Sbh1
Get3 Monomer (kDa)	41.3	38.8	40.2	39.4
Get3 Oligomer (kDa)	82.5	155	161	157
M _w measured (kDa)	80.8	153	198	238
Difference (kDa)	-2.5	-2.0	+37.1	+80.6
TA M _w (kDa)	-	-	8.8	9.3
Difference/TA M _w	-	-	4.2	8.6

SAXS	<i>M</i> γ Get3 (ADP•AlF ₃)	<i>S</i> ϵ Get3/ Ysy6	<i>S</i> ϵ Get3/ Sbh1 47-82
R _g theoretical (Å)	44.2	-	-
R _g Guinier (Å)	47.0	49.8	52.1
R _g GNOM (Å)	47.5	49.6	51.0
D _{max}	165	173	163

Table 1. MALLS and SAXS statistics

Discussion

The details of the GET pathway are rapidly being discovered. So far, Get3 is the best characterized of the components, yet there are still a number of important questions that remain to be answered. Here we solve the structure of an archaeal Get3. The structure is a tetramer with a hydrophobic chamber that we postulate sequesters TA. We demonstrate that archaeal homologues are capable of TA protein binding. Moreover, we show the first structural information of fungal Get3/TA protein complexes, which is consistent with the biochemical data supporting a tetramer model for TA-binding.

The presence of a Get3 homologue in archaea is exciting and suggests a novel membrane protein-targeting pathway in this domain of life. The lack of homologues of other GET components implies that the pathway, if it exists, will be significantly different. Based on structural homology, the fact that archaeal Get3s can bind TA proteins is not surprising; however, it supports the possibility of an archaeal TA targeting pathway. It is also interesting that an archaeal homologue that does not contain the CxxC motif, *TkGet3*, is capable of both oligomerization and TA binding. This motif is essential in fungal Get3 homologues and may suggest an evolutionary path for these proteins. Unlike certain TA proteins (*e.g.* Sec β and SecE), Get3 homologues are not found universally in archaea. This hints that the presumed pathway may not be essential or may be required for specific substrates. The fact that very closely related archaea differ in having a Get3 homologue is a question for further study (*e.g.* *Pyrococcus abyssi* versus *Pyrococcus furiosus* and *Pyrococcus horikoshii*).

Tetramers of Get3 have been seen in a variety of contexts suggesting that this state plays a functional role in the GET pathway. In the initial functional identification of TRC40 (the mammalian homologue of Get3), the protein isolated from in vitro translation in a reticulocyte

lysate was seen in a large complex ranging in size compatible with the tetramer (Stefanovic and Hegde, 2007). Similarly, the crosslinking of a mammalian Get3/TA complex was consistent with tetramers by SDS-PAGE (Favaloro et al., 2008). (Bozkurt et al., 2009) purified a fungal Get3/TA protein complex that they determined was a Get3 tetramer by analytical ultracentrifugation and SEC. This complex, in an *in vitro* insertion assay using mammalian ER microsomes, is competent for membrane insertion. (Favaloro et al., 2010) completed a similar study using a mammalian Get3/TA complex. In this case, the size of the complex was compatible with a tetramer by SEC and was also competent for membrane insertion. We have now replicated these results in a purified fungal system. Thus, in the cases where functional insertion was demonstrated, the oligomeric state of Get3 is a tetramer.

It is clear that eukaryotic Get3 can exist as a stable dimer and it is likely this state plays a functional role. We, and others, have demonstrated that Get3 is a tetramer when bound to a TA substrate and this complex is capable of TA insertion. This conflicts with the dimer model suggested based on earlier structures. As we now present a contrasting model, it is useful to posit the differences in the two models. The dimer model of Get3 TA binding suggests that the HSD captures the length of a TM along a hydrophobic groove with a floor provided by helix 6 and helices 4, 5, 7 and 9 providing the walls. In the unbound transition state structure, the dimensions of the groove seem compatible with a hydrophobic TM helix; however, while covering three sides of the protein the groove leaves one face of the protein exposed to solvent. It has been suggested that helix 8 solves this by becoming ordered upon TA binding to cover the exposed face. Consistent with this, mutations in the walls of the groove and deletion of helix 8 prevent TA binding (Mateja et al., 2009; Yamagata et al., 2010). Similar mutations at the bottom of the groove do not significantly affect TA binding; however, they clearly play a role in the targeting process (Mateja et al., 2009; Suloway et al., 2009). The dimer

model requires that a structured groove accommodate a wide variety of TA substrates along with necessitating the binding of an α -helix. This is different from the model for SRP binding of a hydrophobic signal sequence, which uses a flexible loop to form its helical binding pocket (Bernstein et al., 1989; Keenan et al., 1998).

The tetramer model resolves a number of the issues that arise in the dimer model. Here, the residues in helix 6 that affect nucleotide hydrolysis do not contact the hydrophobic cage directly and would not be expected to directly affect TA-binding. Instead, they would relay the binding of substrate and oligomerization state to the NHD. Moreover, in a deuterium exchange assay the binding of TA leads to exposure of the C-terminus of helix 6, as seen in the tetramer; whereas, in the dimer model this should be occluded by TA (Bozkurt et al., 2009). The binding of a hydrophobic TA stabilizes the tetrameric state of fungal Get3. This suggests that energy must be input to destabilize the complex. It also suggests why a Get3 complex with cytochrome b_5 , a more hydrophilic TA, does not require energy for insertion as it is likely less stable (Stefanovic and Hegde, 2007; Favaloro et al., 2010). It is interesting to note that in the conditions tested here the archaeal tetramer is significantly more stable than eukaryotic homologues independent of bound TA.

Both Get3 and ArsA are members of the ‘deviant Walker A motif’ family of ATPases (Koonin, 1993). ArsA is structurally very similar to Get3 (Suloway et al., 2009) and uses ATP hydrolysis to facilitate arsenite export (Walmsley, 1999). Binding of arsenite stimulates the monomer (effectively a pseudo-dimer) to dimerize (analogous to a Get3 tetramer) (Ching et al., 1991). Structures of ArsA exist only as monomers, so there appears to be an analogous oligomerization in this related system. It would be interesting to see if the solution structure of ArsA bound to arsenite is compatible with the tetramer described here. Other members of

this ATPase family that have been characterized are soluble dimers that also form higher order functional complexes either as homo-oligomers (e.g., Soj; (Leonard et al., 2005) and MinD; (Hayashi et al., 2001)) or as hetero-oligomers (e.g., NifH; (Schindelin et al., 1997)); therefore, a role of higher order oligomerization may be a general feature of this family.

A remaining complication in this study is the stoichiometry of the tetramer/TA complex. The current idea is that a single TA protein binds to the Get3 complex. While this is attractive, there is no biochemical data that supports this; indeed it is difficult to prove. It is clear that our tetramer complex contains minimally a single TA protein; however, the biophysical data suggests that there are more copies bound. The size of the chamber easily accommodates a single TM with ample space for additional TMs. Moreover, the hydrophobic nature of tetramer formation presumably allows for flexibility of the chamber that may expand to bind more substrates. A recent study by (Leznicki et al., 2011) used chemical modification to the TM of the TA to explore the flexibility of the TA binding pocket. Addition of a single large polyethylene glycol (PEG) adduct to the TM did not inhibit binding or insertion; however, modification at two sites prevented binding independent of whether the attachments were on the same or opposite sides of a presumed helix. The single site addition is clearly compatible with a dimer model; however, it is inconsistent with a second binding site on the same side not binding. The tetramer model is also consistent with a single site modification as the two proposed chamber access points could accommodate both protein or extended PEG. A second site would have a harder time being accommodated and would presumably reduce the affinity. Overall, these results point to surprising flexibility in binding to Get3.

The role of nucleotide in TA-targeting remains to be determined. All of the current evidence demonstrates that nucleotide hydrolysis is required at the membrane but not for TA-binding

(Stefanovic and Hegde, 2007; Favaloro et al., 2008; 2010). Mutants deficient in nucleotide binding efficiently bind to TA-substrate in both in vitro (Stefanovic and Hegde, 2007) and heterologous expression assays (Yamagata et al., 2010). Structural studies suggest that Get3 undergoes distinct conformational changes from an ‘open’ to ‘closed’ form that is stimulated by the nucleotide state. The closed form of Get3 has only been seen in the presence of nucleotide, independent of the γ -phosphate (Mateja et al., 2009; Bozkurt et al., 2009). The open form has been seen in both the apo and ADP complex (Mateja et al., 2009; Suloway et al., 2009; Yamagata et al., 2010). This suggests that there is conformational flexibility with bound nucleotide favoring the ‘closed’ state (Chartron et al., 2010). Neither our purified *Mj*Get3 nor our *Sa*Get3/TA complex contain bound nucleotide consistent to what has been seen before (Bozkurt et al., 2009; Favaloro et al., 2010). Purified Get3/TA complexes require nucleotide to stimulate insertion implying that the nucleotide-binding pocket (NBP) is solvent accessible (Bozkurt et al., 2009; Favaloro et al., 2010). With a bound transition state analog the NBP is closed to solvent (Mateja et al., 2009); however, in the closed form bound to ADP the switch II loop has moved, exposing the NBP (Bozkurt et al., 2009). This conformation cannot hydrolyze ATP and is similar to the position of switch II in our *Mj*Get3 tetramer. Therefore, in the Get3 tetramer the NBP is more accessible allowing diffusion in or out of nucleotide. The high cellular concentration of ATP makes it likely that this is the bound form. Disruption of the tetramer would be required for switch II to occupy a hydrolysis competent conformation. This disruption would result in substrate release and may be facilitated by factors at the membrane.

In the fungal GET pathway each of the soluble proteins exists minimally as dimers. Sgt2, the first protein in the pathway to specifically bind the TA (Tobaben et al., 2003; Wang et al., 2010), is a homodimer with an N-terminal dimerization domain (Liou and Wang, 2005). Get4

and Get5, which link Sgt2 to Get3, are hetero-tetramers with a C-terminal dimerization domain in Get5 (Chartron et al., 2010). This all suggests a larger functional complex in TA protein recognition and delivery (Chartron et al., 2011). The ability of Get4, minimally present in two copies, to bind directly to a dimer of Get3 is consistent with the possible specific recognition of a tetramer or two dimers.

All of this allows us to suggest a modified model for Get3 mediated targeting of a TA protein (Figure 7). (1) Get3 in its apo form is a stable and soluble dimer in equilibrium between an ‘open’ and ‘closed’ form. (2) Binding of nucleotide shifts the equilibrium towards the ‘closed’ form that is compatible with binding Get4. This is now competent for TA binding. (3) Binding of the TA results in tetramer formation. Conformational changes to form the tetramer cause release from Get4 and diffusion of the Get3/TA complex to the membrane. (4) Here, the receptor complex stimulates release of TA and insertion into the membrane. Get3 now returns to the dimer state to bind new TA substrates.

A related structural study on archaeal Get3 was published while this manuscript was in review. Their crystal structure of Get3 from the archaea *Methanothermobacter thermautotrophicus* (*MtGet3*) is consistent with the results presented here in that the archaeal homologue is structurally similar to fungal counterparts (Sherrill et al., 2011). In this case, a dimer was specifically purified and crystallized; therefore, they do not investigate the tetramer that we report. Interestingly, the loops that extend to form our tetramer are also extended in the dimer of *MtGet3* perhaps explaining why the archaeal tetramer is more stable. Excitingly, there they were able to demonstrate that *MtGet3* can facilitate TA insertion using a protease protection assay. This difference from our result might be assay specific or could be related to the species of Get3 and TA protein tested.

Also while this manuscript was in review, two reports on the interaction of Get3 with the membrane proteins Get1 and Get2 were published (Mariappan et al., 2011; Stefer et al., 2011). In both studies, structures of Get3 dimers are bound to the soluble domains of either Get1 or Get2. The structures suggest how Get1 and Get2 can recognize Get3 and facilitate release of a TA protein. These structures do not contain TA proteins. We believe the evidence is compatible with a model in which a tetrameric Get3/TA protein complex is captured by Get2 then disrupted by Get1 to release the TA protein for insertion.

Get3 is a dynamic protein that undergoes a complex series of conformational changes in delivery of a TA protein to the ER. Here we present the first structural information of Get3/TA complex from a heterologously expressed system demonstrating that Get3 in this state is a tetramer. The tetramer model suggests a TA is sequestered within a hydrophobic chamber. Further studies are required to establish the role of the tetramer in vivo. An unresolved point is the stoichiometry of the TA proteins to Get3, which probably requires other GET components to determine. Finally, the functional role of the Get3 archaeal homologue is a tantalizing question, particularly with the broader context of the detailed molecular mechanism of TA protein targeting by the GET pathway.

Materials and Methods

Cloning, expression and purification

All Get3 homologues were amplified from genomic DNA and cloned into a pET33b vector (Novagen) modified to contain only an N-terminal 6×His tag. Genomic DNA for *Mj*Get3 (MJ_1142) from *M. jannaschii* DSM 2661 (ATCC), *Tk*Get3 (TK_0994) from *T. kodakaraensis* KW128 (Santangelo et al., 2007), *Mm*Get3 (MmarC7_1163) from *M. maripaludis* C7 (ATCC), and *Sl*Get3 from a previous study (Suloway et al., 2009). For *Mj*Get3, site-directed mutagenesis was used to generate a truncation encoding amino acids 12–349 of *Mj*Get3 (*Mj*Get312–349) and 12–333 of *Mj*Get3 (*Mj*Get312–333). For co-expression, *Mj*Get312–349, *Tk*Get3 and *Sl*Get3 were cloned into the first multiple-cloning site (MCS) of pACYCDuet (Novagen). TA proteins (Ysy6 YBR162W-A, Sbh1 YER087C-B, Secβ (Kinch et al., 2002; Van Den Berg et al., 2004), SecE MJ_0371, MtrA MJ_0851, MtrB MJ_0850) were amplified from genomic DNA and cloned into pMAL-C2 (NEB) modified to contain a thrombin site between MBP and the MCS. Truncations of Sbh1 were generated by site-directed mutagenesis. Gene annotations are from KEGG (<http://www.genome.jp/kegg/>).

Get3 homologues were expressed in BL21-Gold(DE3) (Stratagene) in 2×YT at 37°C for 3h (induced at $A_{600}=0.6$ with 0.3mM isopropyl-β-D-thiogalactoside). Cells were pelleted, resuspended in Buffer A (50mM Tris pH 7.5, 300mM NaCl, 10mM β-mercaptoethanol (βME)) with protease inhibitors, and lysed through a ML-110 microfluidizer (Microfluidics). Lysate was centrifuged and supernatant was passed over Ni-NTA resin (Qiagen), washed with Buffer A with 10mM imidazole and eluted in Buffer A with 200mM imidazole. The eluate was incubated with 2U thrombin per ml (Sigma) at room temperature (RT) while dialyzing against Buffer A for 16h. The reaction was stopped with 1mM PMSF and passed over Ni-NTA resin

to remove uncleaved product and contaminants. Flow-through was purified on a Superdex 200 column (GE Healthcare) (10mM Tris pH 7.5, 100mM NaCl, 10mM β ME). Selenomethione derivatives were expressed by auto-induction media and purified the same method as native (Studier, 2005). *Sa*Get3 tetramers were analyzed for extraneous peptides by tryptic digest followed by LC/MS at the Caltech Protein/Peptide MicroAnalytical Laboratory.

Crystallization

*Mj*Get3₁₂₋₃₄₉ crystallized in the P2₁ form in 2 days at RT by sitting-drop vapor diffusion by mixing 1 μ l of 10mg/ml *Mj*Get3 (10mM Tris pH 7.5, 100mM NaCl, 10mM β ME, 2mM MgCl₂, 2mM ADP, 0.5mM AlCl₃, 8mM NaF) with 1 μ l of reservoir solution (0.1M Na₂SO₄ and 9% (w/v) PEG 3350). Crystals were cryo-protected with artificial mother liquor containing 20% glycerol or 17.5% sucrose and 17.5% xylitol before flash freezing in liquid N₂. Seleno-methione crystals were obtained in the same way. *Mj*Get3₁₂₋₃₃₃ crystallized in the P2 form after 1 day in drops of 1 μ l of 10mg/ml *Mj*Get3 (10mM Tris pH 7.5, 100mM NaCl, 10mM β ME, 2mM MgCl₂, 2mM ADP) and 1 μ l 0.2M Na₂SO₄ and 10% (w/v) PEG 3350 and cryoprotected with 20% ethylene glycol.

Data collection, structure solution and refinement

All native data collection was done at SSRL BL12-2 at a wavelength of 1.000Å at 100K. Selenomethionine derivative data were collected at the APS GM/CA-CAT BL23ID-D at a wavelength of 0.9795Å at 100K. Diffraction data were integrated with iMosflm (Leslie, 1992) and scaled with CCP4/Scala for P2₁ data (CCP4, 1994) and XDS for P2 data (Kabsch, 2010). Molecular replacement (MR) with native P2₁ data was performed with CCP4/Phaser (McCoy, 2007) with a NHD from 3IBG. MR for native P2 data was by phenix.automr used a starting model of a NHD monomer from *Mj*Get3 P2₁. The P2₁ form consisted of a single *Mj*Get3

tetramer in the asymmetric unit. The P2 form contained four monomers assembled in two crystallographic tetramers aligned along the long axis of symmetry. Rounds of model building and refinement were done with Coot (Emsley and Cowtan, 2004) and phenix.refine (Adams et al., 2002). Global non-crystallographic symmetry (NCS) was used in the P2₁ refinement with weights calculated in phenix.refine. Residues that we were not able to resolve in both crystal forms varied between subunits but included the N-terminus (2-23/25), the loop between helices 8 and 9 (202-209) and the C-terminus (333/334-349). In P2, we could see density in chain A that is consistent with helix 1 of chain A in P2₁ but were unable to build into it with confidence. Additionally, we were unable to convincingly model density in the nucleotide-binding pocket near the magnesium and aluminum fluoride-binding site. As we see it in both forms it could be a sulfate ion; however, we do not have direct evidence for this. TLS (translation/libration/screw) vibrational motions were calculated using the TLSMD server (Painter and Merritt, 2006a; 2006b) and used in the refinement. After initial modeling and refinement the P2 model then refined against data corrected by the Diffraction Anisotropy Server (Strong et al., 2006) limiting the resolution in directions a^* to 3.3Å, b^* to 2.9Å and c^* to 3.4Å. The final P2₁ model had an R_{work} of 25.1% and an R_{free} of 28.6% with residues in the Ramachandran plot in 97.5% preferred, 2.5% allowed and 0.0% in the disallowed and restricted regions. The final P2 model had an R_{work} of 27.0% and an R_{free} of 29.6% with residues in the Ramachandran plot in 96.2% preferred, 3.8% allowed and 0.0% in the disallowed and restricted regions. Ramachandran statistics are taken from PHENIX. All structure figures were made using PyMOL (Schrödinger, LLC) except for 2a, 5, and S5 which were made using UCSF Chimera (Pettersen et al., 2004).

Pull-downs

*Sc*Get3, *Mj*Get3 or *Tk*Get3 and TA proteins were co-expressed in BL21-Gold(DE3) *E. coli*. Soluble complexes were purified in two steps using amylose resin (NEB) and Ni-NTA resin. *Sc*Get3 complexes with Sbh1-truncations were purified by the same method followed by incubation with 2 U of thrombin per ml at room temperature overnight.

Microsome insertion assay

The *Sc*Get3 complex with MBP-Sbh1-op was co-expressed and purified as for the pull-downs followed by SEC on a Superdex 200 10/300. MBP-Sbh1-op was purified using amylose resin. Microsomes from WT and $\Delta get3$ strains were prepared as in (Schuldiner et al., 2008). Purified complex or MBP-Sbh1-op and $\Delta get3$ microsomes were used for the insertion assay using the conditions reported in (Bozkurt et al., 2009).

S. cerevisiae translation extracts were prepared essentially as in (Wu et al., 2007), and included an additional centrifugation step at 49,000 rpm in an Sw55Ti for 30 minutes after the low speed centrifugation step (following cell lysis). In vitro translations were carried out as in (Wu et al., 2007). Translation reactions were performed with $\Delta get3$ extracts in a 10 μ L scale with 10 μ Ci [35S] methionine in the presence of recombinant Get3 (concentrations indicated on gel). To assay post-translational TA protein insertion, the following was added after 30 minutes: 1 mM cycloheximide, 1 μ L energy mix (8.3 mM ATP, 1,7 mM GTP, 200 mM creatine phosphate, 600 mM KOAc, 10 mM MgOAc), and 0.006 U/ μ L YRMs (WT or $\Delta get3$).

SEC-MALLS

Purified proteins were run on a Shodex KW-804 column (10mM Tris pH 7.5, 100mM NaCl, 10mM β ME) with MALLS data collected on a DAWN HELEOS and Optilab rEX detector (Wyatt). Data were processed using ASTRA (Wyatt) software.

SAXS

Purified *Mj*Get3₁₂₋₃₄₉ and *Sc*Get3 complexes with TA-substrates were dialysed against 20mM Tris pH 7.5, 250mM NaCl and 10mM β ME. *Mj*Get3₁₂₋₃₄₉ samples were also prepared by dialysis against the same buffer containing 2mM MgCl₂, 2mM ADP, 8mM NaF, 0.5mM AlCl₃ and 1mM ZnSO₄. Solution small-angle X-ray scattering (SAXS) experiments were done at SSRL BL4-2 at RT. SAXS diffraction images were processed using SASTool and PRIMUS (Konarev et al., 2003), data were analyzed with PRIMUS/autorg/autoprod, particle distance functions were generated with GNOM/autognoM (Svergun, 1992) and ab initio shape determination was done with DAMMIN (Svergun, 1999).

ICP-MS measurement

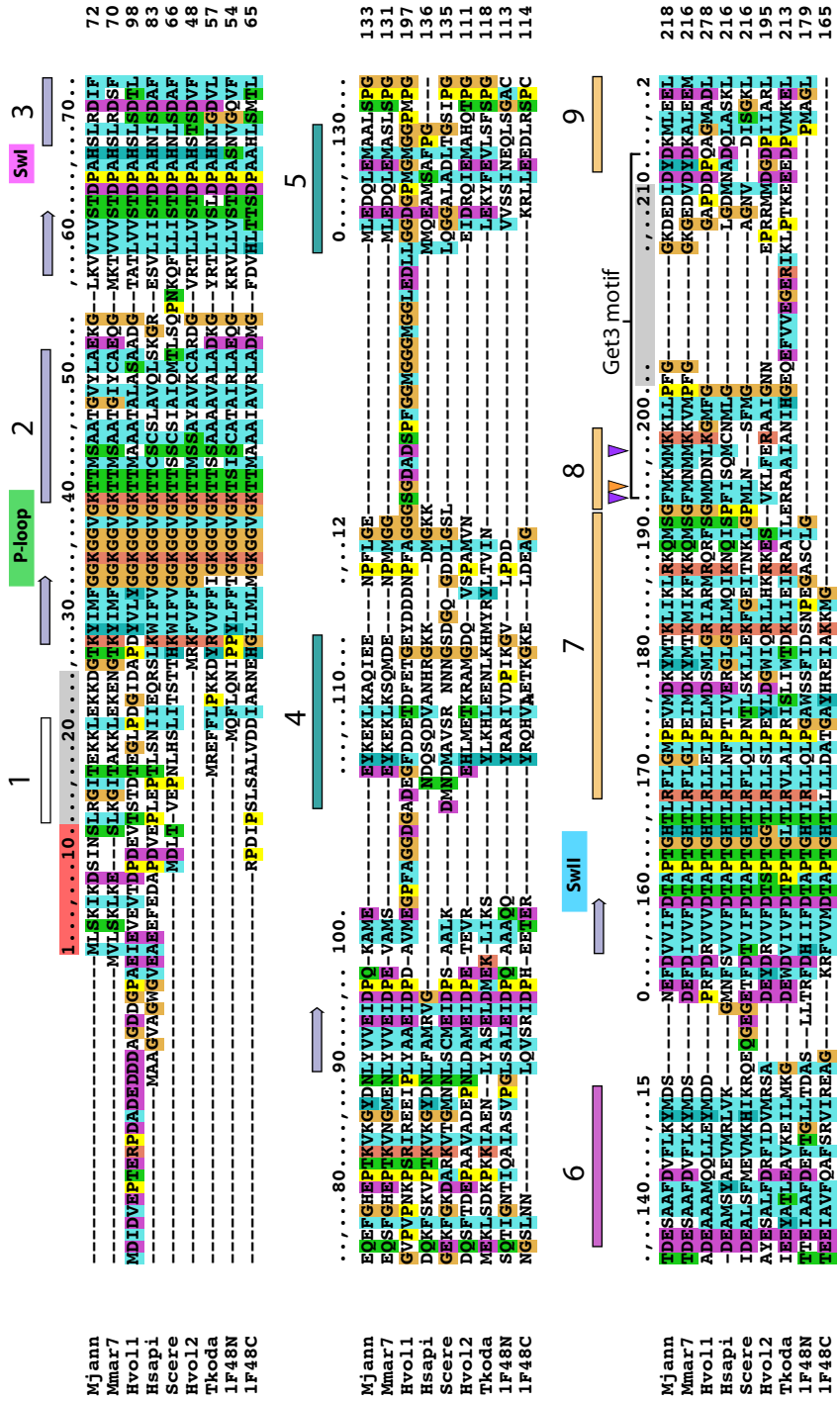
The zinc and magnesium occupancy was quantified by ICP-MS similar to (Yamagata et al., 2010). Samples were measured using a HP-4500 ICP-MS (Agilent Technologies). The concentration of the Get3/TA complex was measured at between 3.24 and 3.48 μ M assuming a Get3 tetramer bound to either one or four substrates. The zinc concentration was measured at 5.66 μ M consistent with two ions per tetramer. The magnesium concentration was measured at 1.9 μ M indicating less than one ion per tetramer, consistent with no detectable amount of magnesium binding, which is required for nucleotide binding.

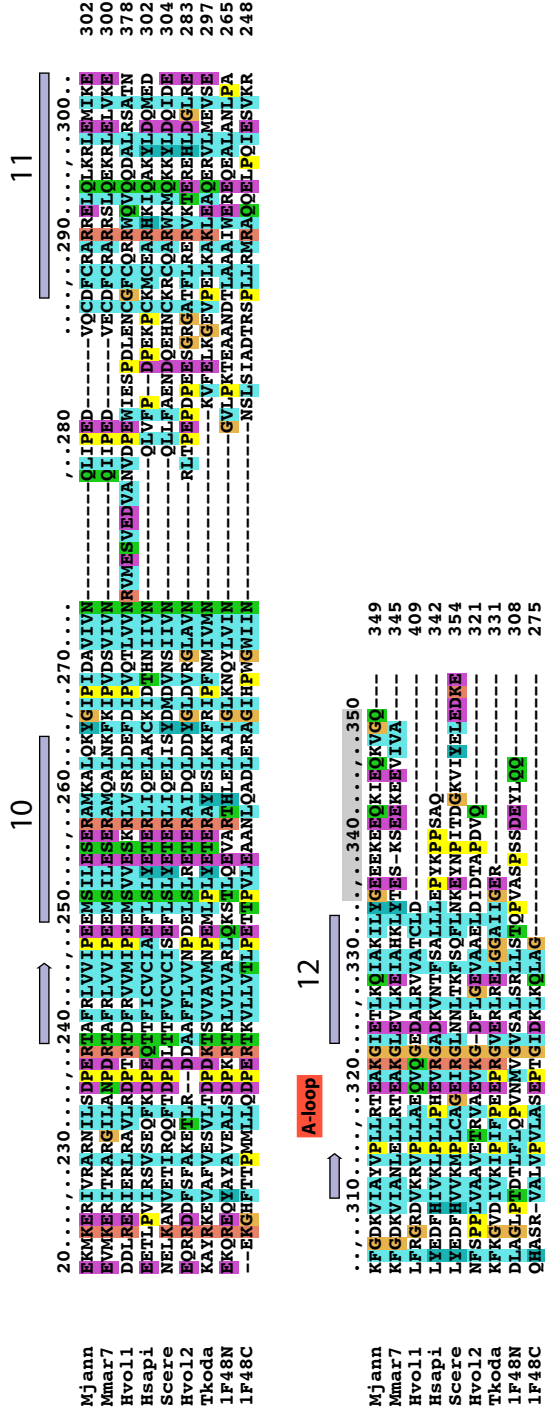
ACKNOWLEDGEMENTS

We thank Grace Wu for help generating the TA protein constructs and Shu-ou Shan and Raymond Liu for help in microsome preparation. We thank DC Rees, JW Chartron, S Tanaka, HB Gristick and SO Shan for critical reading of the manuscript. We thank members of the laboratory for support and useful discussions. *T. kodakaraensis* genomic DNA was kindly provided by Tom Santangelo (Ohio State). We thank Graeme Card, Ana Gonzalez and Michael Soltice for help with data collection at SSRL BL12-2, Justin Chartron, Michael Becker and Craig Ogata for help with data collection at APS GM/ CA CAT BL23ID-D, and Tsutomu Matsui and Hiro Tsuruta for help with data collection and processing at the bioSAXS SSRL BL4-2. ICP-MS instrumentation was provided by the Environmental Analysis Center at the Caltech and we were aided by Nathan Dalleska. We are grateful to Gordon and Betty Moore for support of the Molecular Observatory at Caltech. Operations at SSRL and APS are supported by the US DOE and NIH. This work is funded by grants to WMC including, the Searle Scholar program, a Burroughs-Wellcome Fund Career Award for the Biological Sciences and a National Institutes of Health Grant R01 GM097572. The atomic coordinates and structure factors have been deposited in the RCSB Protein Data Bank, <http://www.pdb.org> (PDB ID codes 3UG6 and 3UG7 for the P2₁ and P2 forms, respectively).

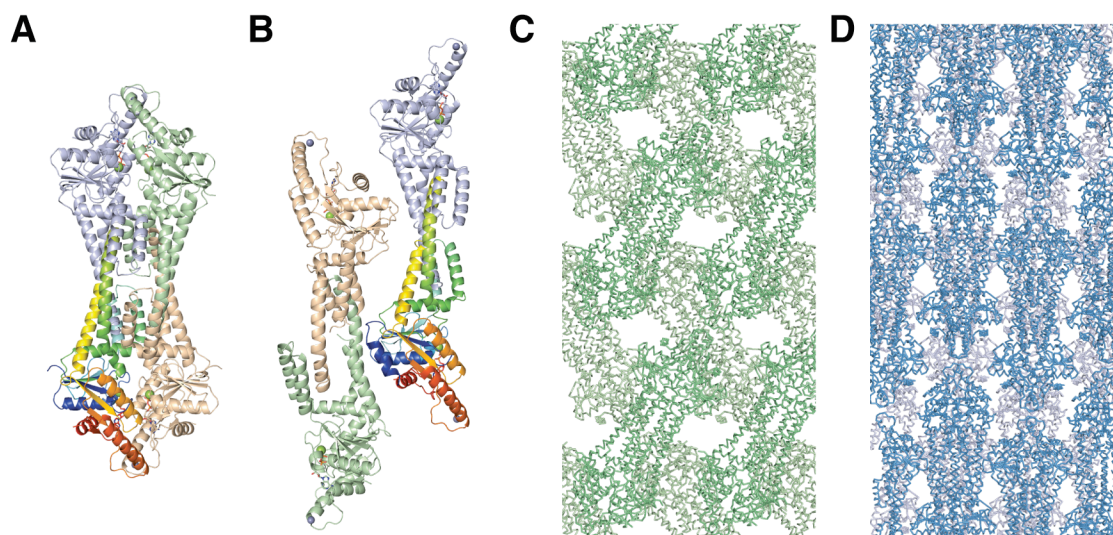
Author contributions: CJMS performed all of the experiments except the in-vitro translation insertion assay. MER provided microsome membranes and performed the in-vitro translation insertion assay. CJMS and WMC designed experiments and wrote the manuscript.

Supplementary Figures



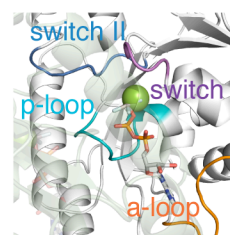
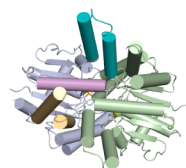
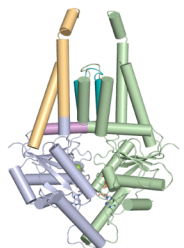


Supplementary Figure 1. Sequence alignment of Get3 Sequences were aligned with ClustalX (<http://www.clustal.org>). *M. jannaschii* (Mjann), *M. maripaludis* C7 (Mmar7), *Haloferax volcanii* arsA1 (Hvol1), *Homo sapiens* (Hsapi), *S. cerevisiae* (Scere), *H. volcanii* arsA2 (Hvol2), *T. kodakaraensis* (Tkoda), the N- and C-terminal sequence of *E. coli* Arsa (1F48N and C). Amino acids are colored according to conservation and type from ClustalX output. Features are labeled and colored according to Figure 1D, helices are numbered and secondary structure is colored as in Figure 1E. Sequence numbering is for *Mj*Get3. Truncated residues for *Mj*Get3 construct highlighted in red and disordered polypeptide in gray. Mutants tested are marked by triangles colored as in Figure 5B.

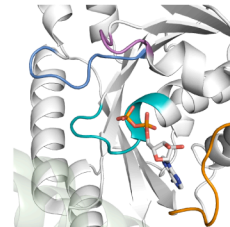
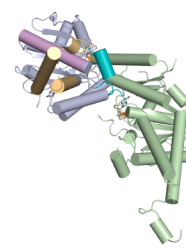
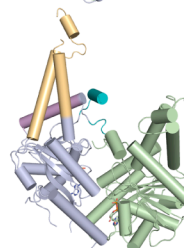


Supplementary Figure 2. Asymmetric units of the two *MjGet3* crystal forms. (A) *MjGet3* P2₁ crystal form colored as in Figure 1A. (B) As in (A) for the P2 crystal form. Note that the tetramer is formed on a crystallographic axis. (C) Crystal packing of the *MjGet3* P2₁ crystal form (green). (D) As in (C) for the P2 crystal form (blue).

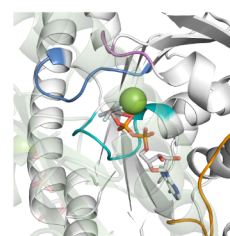
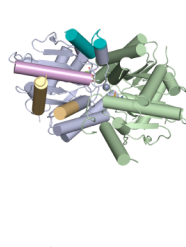
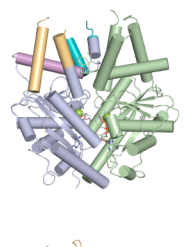
3UG6
Suloway *et al.* 2011
M. jannaschii
ADP
closed



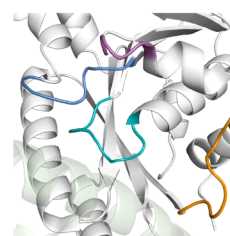
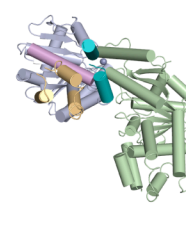
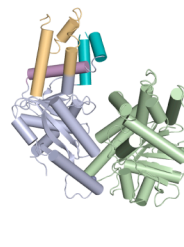
3IBG
Suloway *et al.* 2009
A. fumigatus
ADP
open



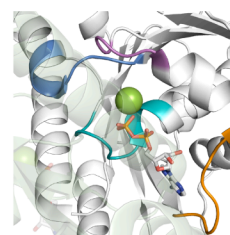
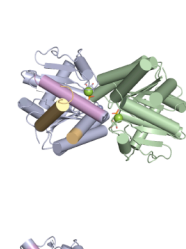
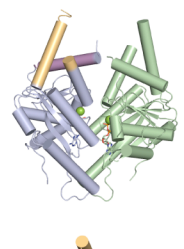
2WOJ
Mateja *et al.* 2009
S. cerevisiae
ADP•AIF₄
closed



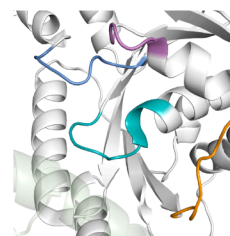
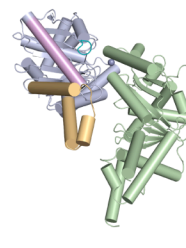
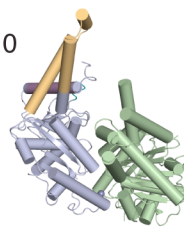
2WOO
Mateja *et al.* 2009
S. pombe
apo
open



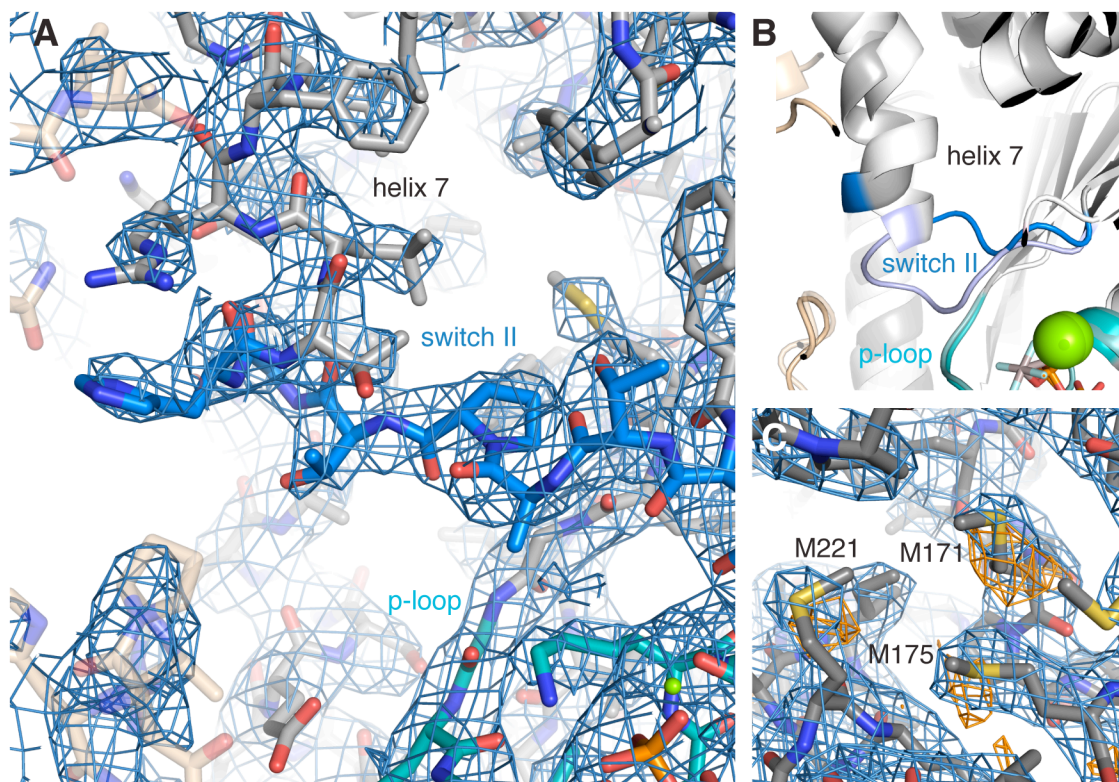
3IQX
Bozkurt *et al.* 2009
C. thermophilum
ADP
closed



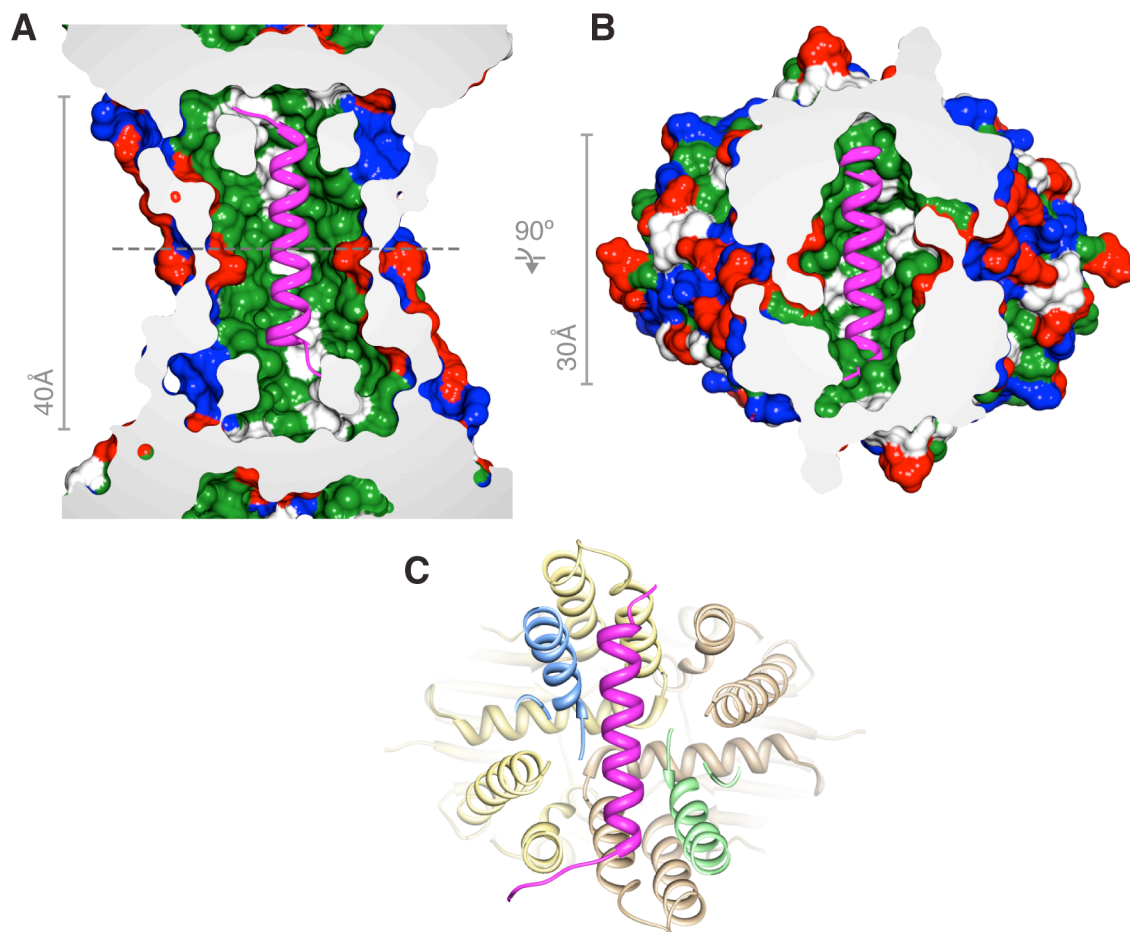
3A36
Yamagata *et al.* 2010
S. cerevisiae
apo
open



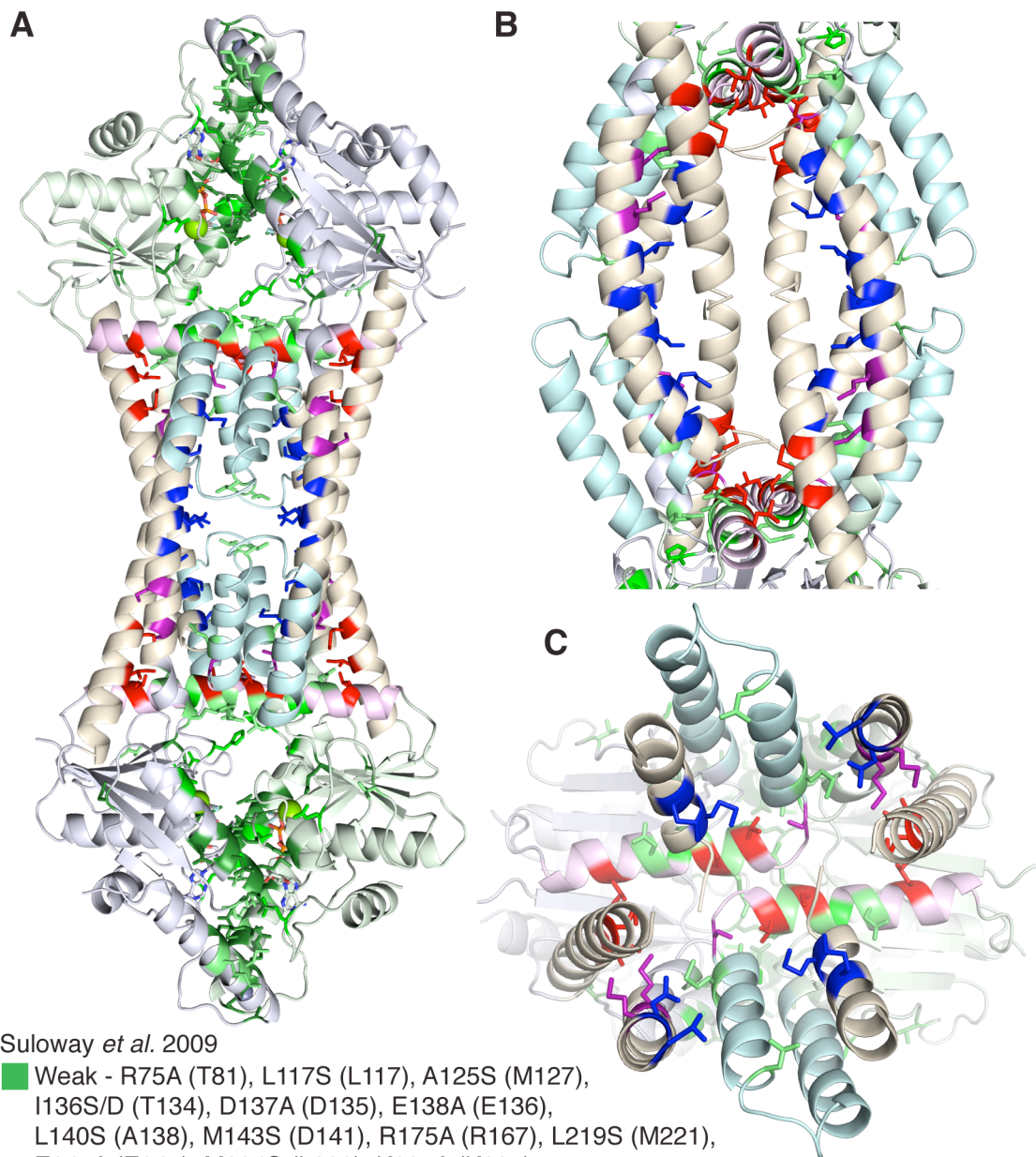
Supplementary Figure 3. Comparison of Get3 structures. A sampling of existing Get3 structures are compared. Each structure has been aligned to the lavender *Mj*Get3 monomer. The first column lists the PDBID, publication, species, nucleotide and open or closed state associated with the structure. The next two columns show a view as in Figure 1E and a view as in Figure 2B each colored as in Figure 1E. The last column shows the nucleotide hydrolase domain colored as in Figure 1D with all structures aligned to the P-loop.



Supplementary Figure 4. *MjGet3* switch II and electron density. (A) The *MjGet3* switch II loop shown in stick representation colored as in Figure 1D with electron density contoured at 1.5σ (blue mesh). (B) A cartoon representation comparison of the *MjGet3* switch II loop to that of *ScGet3* (2WOJ). Both structures are colored as in Figure 1D except 2WOJ has been lightened. (C) A representative view of an anomalous difference map calculated for the Se-met dataset (contoured at 2.8σ , orange mesh) to ensure registry for the cage (methionines are labeled).



Supplementary Figure 5. Sec β fits inside the tetramer cavity. (A) Cartoon representation of Sec β from 1RH5 (magenta) aligned along the long axis inside the tetrameric *Mj*Get3 cavity similar to Figure 2A. General dimensions are noted (B). Second view of the cavity rotated 90° from (A) with Sec β rotated to be orthogonal to the long axis. (C) Same orientation as (B) in cartoon representation.



Suloway *et al.* 2009

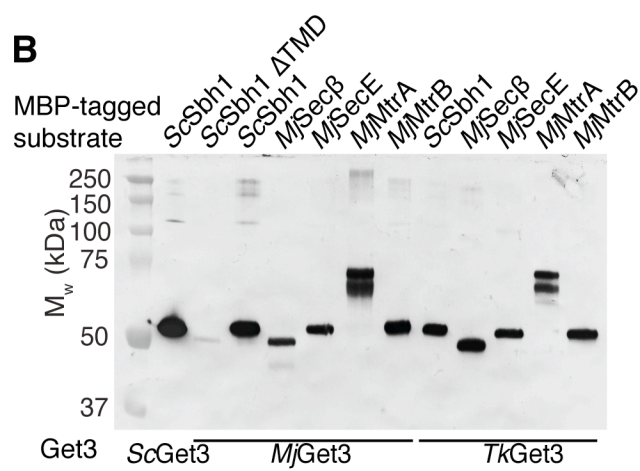
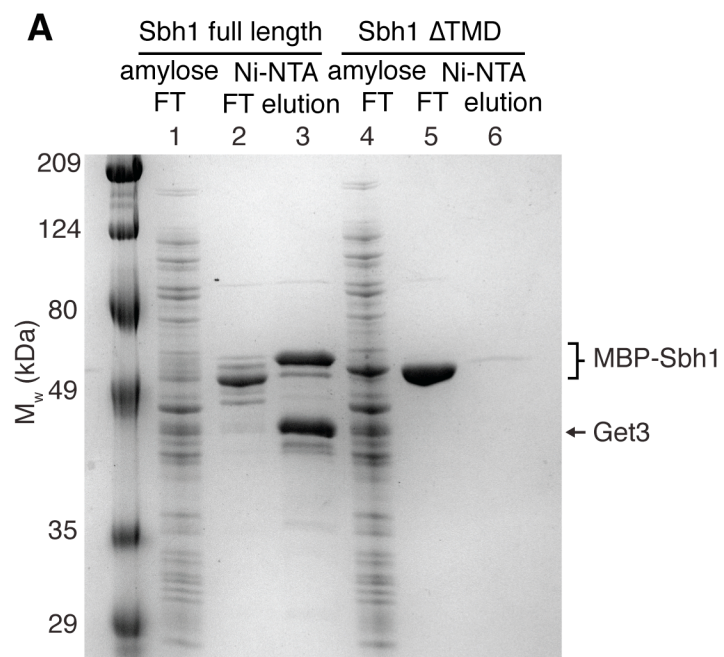
- Weak - R75A (T81), L117S (L117), A125S (M127), I136S/D (T134), D137A (D135), E138A (E136), L140S (A138), M143S (D141), R175A (R167), L219S (M221), E245A (E247), M294S (L292), K297A (K295)
- Moderate - K69A (E75), S141A (A139), H172A (H164), S248A (S250), E258A (A260), E320A (E318)
- Strong - K26A/R (K34), G30R (G38), D64A (D70), E87A (E93), F246A (E248), L247S (M249), Y250A (L252), E251A (E253), E253A (E255), D265A (P267), C285/288T (C283/C286), R291A (R289), Y298A (R296)

Mateja *et al.* 2009

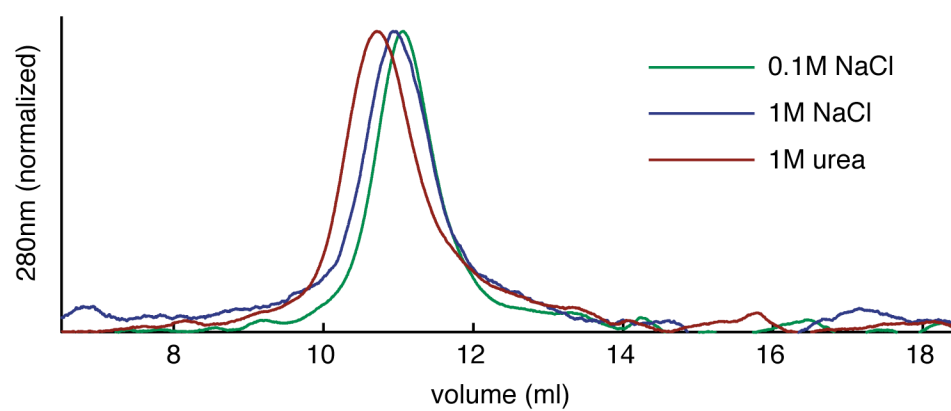
- ATPase - I136D (T134), A139D (S137), M146D (L144), L219D (M221), V223D (I225)
- TA-protein binding - F190D (L182), I193D (L185), L201D (M193), F204D (M196)
- Both - I133D (S131), L183D (M175), L187D (M179)

Supplementary Figure 6. Get3 mutational analysis. Three additional views of Figure 2D.

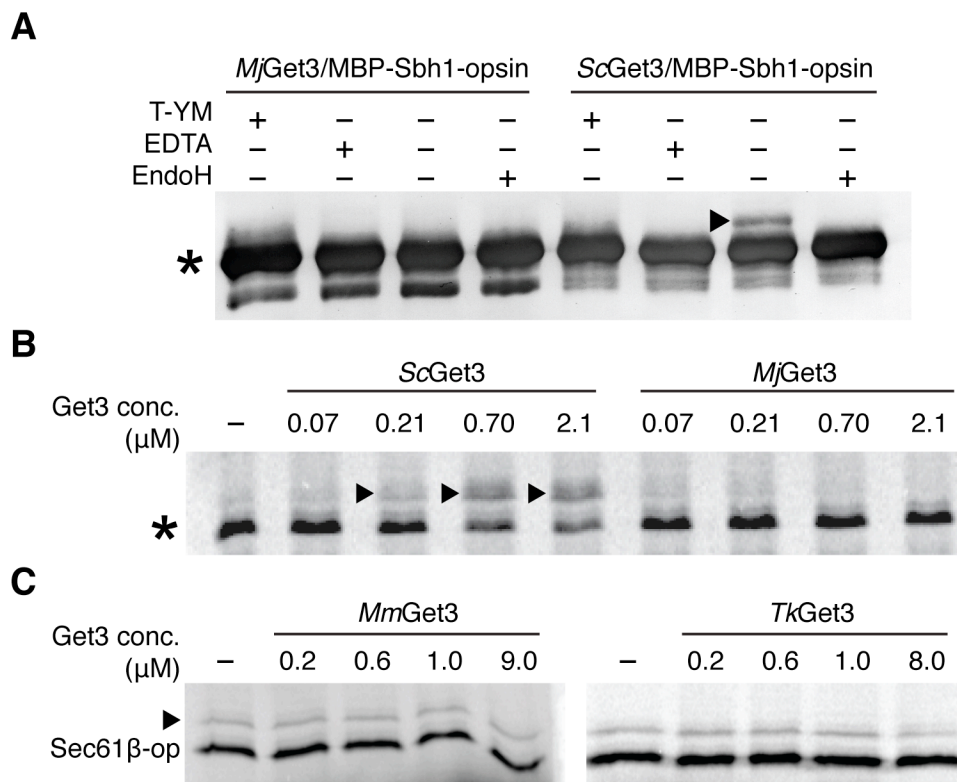
Colors and mutants listed for *Ss*Get3 followed by the equivalent *Mj*Get3 residue in parenthesis. (A) Full view with the same orientation as Figure 1A. (B) Rotated approximately 90° from (A) viewed along the putative groove. (C) Inside the cage looking towards a dimer (same orientation as Figure 2B).



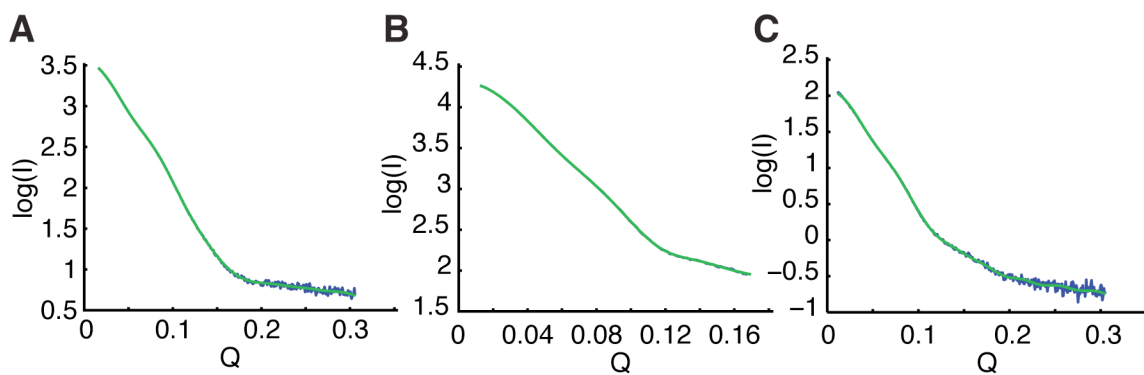
Supplementary Figure 7. Controls for the purification of complex from various species. (A) SDS-PAGE visualized with Coomassie staining demonstrating the two-step affinity purification of the complex of *Mj*Get3 and MBP-Sbh1 and *Mj*Get3 and MBP-Sbh1 without the TM of Sbh1 (Sbh1 Δ TMD). Lanes 1 & 4 are the flow through for the amylose affinity step containing impurities and uncomplexed Get3. Lanes 2 & 5 are the flow through for the Ni affinity step using the elution from the amylose column containing uncomplexed MBP-Sbh1. Lanes 3 & 6 contain the elution from Ni affinity column with either bands for the MBP-SBh1 and Get3 eluted as a complex for the full length Sbh1 or no bands for the absence of a complex for Sbh1 Δ TMD. (B) A Western blot against the MBP tag on the MBP fusions to various TA proteins when complexed to *Mj*Get3 and *Tk*Get3 as shown in Figure 3B with the addition of Sbh1 Δ TMD.



Supplementary Figure 8. SEC of *MjGet3* in high salt and denaturant. Size-exclusion chromatogram of *MjGet3* run in a control buffer of 0.1M NaCl (green), high salt 1M NaCl buffer (blue) and a buffer contain 1M urea as a denturant (red).



Supplementary Figure 9. Membrane insertion assays. (A) Membrane insertion assay of MBP-Sbh1-op by purified *Mj*Get3/MBP-Sbh1-op and *Sc*Get3/MBP-Sbh1 as in Figure 5A Lanes 1–4. In vitro translation of a N-terminally truncated MBP-Sbh1-op in the presence of (B) *Sc*Get3 or *Mj*Get3 into Δ Get3 *S. cerevisiae* microsomes or (C) in the presence of *Mm*Get3 or *Tk*Get3 into wild-type *S. cerevisiae* microsomes analyzed by SDS-PAGE followed by autoradiography.



Supplementary Figure 10. bioSAXS data analysis. Experimental scattering curves (blue) and the fit from best single calculated envelope (green) for (A) *Mj*Get3, (B) *Sl*Get3/Ysy6 complex and (C) *Sl*Get3/Sbh1₄₇₋₈₂ complex.

	<i>Mj</i> Get3 P2	<i>Mj</i> Get3 P2 ₁	<i>Mj</i> Get3 P2 ₁ Se-Met
Data collection			
Space group	P2	P2 ₁	P2 ₁
Cell dimensions			
<i>a</i> , <i>b</i> , <i>c</i> (Å)	71.95, 149.64, 72.89	85.74, 76.87, 127.81	86.71, 76.54, 128.12
α , β , γ (°)	90.00, 94.24, 90.00	90.00, 108.83, 90.00	90.00, 108.82, 90.00
Resolution (Å)	37.41-3.19 (3.36-3.19)*	50.00-3.30 (3.48-3.30)*	50.00-3.80 (4.01-3.80)
<i>R</i> _{merge}	9.5 (69.0)	15.1 (71.2)	26.4 (70.4)
<i>I</i> / σI	8.5 (1.3)	6.3 (2.1)	8.3 (3.4)
Completeness (%)	95.3 (92.1)	96.2 (95.0)	99.9 (100.0)
Redundancy	2.9 (2.8)	3.1 (2.9)	8.6 (8.7)
Refinement			
Resolution (Å)	37.4-3.2	42.9-3.3	
No. reflections	46,076	22,946	
<i>R</i> _{work} / <i>R</i> _{free}	25.0/29.6	24.2/28.2	
No. atoms	9,906	9,754	
Protein	9,790	9,640	
Ligand/ion	116	114	
Water	0	0	
<i>B</i> -factors	107.7	81.9	
Protein	107.9	82.2	
Ligand/ion	89.3	56.5	
Water	-	-	
R.m.s. deviations			
Bond lengths (Å)	0.018	0.029	
Bond angles (°)	1.60	1.70	
Values in parentheses are for highest-resolution shell.			

Supplementary Table 1. Data collection and refinement statistics

CONCLUSION

In the five years following the discovery of Get3 as a TA protein translocation factor (Stefanovic and Hegde, 2007; Favaloro et al., 2008) there have been rapid advances in our knowledge about how TA proteins are targeted to the ER membrane. The elucidation of the pathway in *S. cerevisiae* lead to the discovery of the membrane receptor Get1/2 (Schuldiner et al., 2008) and upstream factors Get4/5 (Jonikas et al., 2009) putting Get3 in a larger context as part of a group of proteins that form a TA protein translocation pathway.

Structural studies of Get3 demonstrated how an open form of homodimeric Get3 would bind ATP to form a closed form presenting a binding pocket for TA proteins (Hu et al., 2009; Mateja et al., 2009; Suloway et al., 2009; Bozkurt et al., 2009; Yamagata et al., 2010). This was followed by studies showing the structure of Get4 with the N-terminal domain of Get5 forms a binding surface for Get3 and that a dimer of Get5 binds monomers of Get4 at each end to form a heterotetramer (Bozkurt et al., 2010; Chang et al., 2010; Chartron et al., 2010).

Sgt2 was identified as another component of the GET pathway through interactions with Get4/5 (Copic et al., 2009; Chang et al., 2010; Costanzo et al., 2010). Get4/5 was shown to mediate the loading of TA proteins from Sgt2 to Get3 (Wang et al., 2010). The TA protein loading complex in mammals was discovered to be SGTA and BAG6/TRC35/UBL4A analogous to Sgt2 and Get4/5, respectively (Leznicki et al., 2010; Mariappan et al., 2010). The mammalian homolog of Get1 was identified as WRB and shown to mediate TA protein insertion into membranes (Schuldiner et al., 2008; Vilardi et al., 2011). Archaeal homologs of Get3 were found bioinformatically (Borgese and Righi, 2010) and shown to be structurally similar as well as capable of binding and integrating TA proteins into the membrane (Sherrill

et al., 2011; Suloway et al., 2012). This expanded the scope of Get3 from specifically targeting TA proteins to the ER in eukaryotes to include prokaryotic TA protein translocation.

Structures of Get3 complexed with either the soluble domains of Get1 or Get2 along with biochemical evidence has formed the basis for a model for membrane insertion (Mariappan et al., 2011; Stefer et al., 2011). In this model Get2 tethers Get3 complexed with TA protein to the membrane followed by Get1 displacing Get2 and stimulating the release of TA proteins for membrane integration. The larger picture of the structure and function of the sorting complex of Get4/5/Sgt2 along with chaperones has been elucidated by biochemical studies and structures of Sgt2 and the Get5 dimerization domain (Chartron et al., 2011; 2012). Structural evidence from SAXS and biochemistry have elaborated on the nature of the interaction of Get3 with Get4/5 (Chang et al., 2012).

This work has presented a model for TA protein targeting by Get3. The structure of Get3 from *Aspergillus fumigatus* complexed with ADP shows a that dimer of Get3 formed by interactions between the nucleotide hydrolase domains has a similar architecture to members of the SIMIBI class nucleotide hydrolyases NifH and ArsA. Helices that extend out from the NHD contain conserved hydrophobic residues make a plausible site for TA protein binding. By modeling a rotation between the two domains based on the conformational changes between the open and closed form of NifH the dimer interface is brought together along with the putative TA protein binding site. Using a phenotypic rescue assay, the importance of residues at the dimer interface and hydrophobic residues for possible TA protein binding was demonstrated. A model was posited where an open form of Get3 in apo or ADP bound state was transformed into a closed form competent for TA protein binding when ATP was bound, and the TA protein would be released after nucleotide hydrolysis.

A recombinant co-expression system was developed for obtaining a variety of complexes of Get3 with TA proteins for structural studies. Purification by two-step affinity chromatography yielded purified complexes that by SEC showed a higher oligomeric state for Get3 when complexed with TA proteins. Crystals from purified complexes failed to diffract X-rays to sufficient resolution for structural studies. To form complexes with different nucleotide states, an in vitro reconstitution system was developed where solubilized TA proteins were added to Get3 incubated with nucleotide and nucleotide state analogs. Quantities of complex necessary for structural studies have yet to be obtained.

The structure of an archaeal homolog of Get3 from *M. jannaschii* revealed the same basic structural features as fungal Get3. Archaeal Get3 can bind fungal and archaeal TA proteins to form stable complexes indicating its role in TA protein translocation. The *Mj*Get3 structure is a tetramer formed by two Get3 dimers interacting between the putative TA protein binding helices extending out from the NHD. The solution state of *Mj*Get3 and other archaeal Get3 homologs is also tetrameric and can be disrupted by detergent or mutations to the tetramerization interface seen in the structure. Strikingly, at the center of the *Mj*Get3 tetramer is a hydrophobic chamber that accommodates the size and biochemical properties of a TA. Get3 sequesters only the TA as shown by a specific protease protection assay. Fungal Get3 complexed with TA proteins is tetrameric in solution and the idea that Get3 binds a TA protein as a tetramer in a central hydrophobic chamber reconciles previous mutagenesis experiments. Recombinantly co-expressed tetrameric Get3-TA protein complexes are functional, capable of integrating TA proteins into the membrane using in vitro translocation assays. Finally, SAXS shows the shape of tetrameric *Mj*Get3 and Get3-TA protein complexes in solution are very similar and fit the crystal structure of *Mj*Get3. This evidence leads to a

model where tetrameric Get3 sequesters TA proteins in a central hydrophobic chamber, in contrast to previous dimeric binding models.

In recent times our understanding of the GET pathway has developed significantly and given insight into the overall process of integral membrane protein translocation, but there is still much to be explored. The architecture and dynamics of the sorting complexes in yeast and mammals and how TA proteins destined for different locations are differentiated continues to be elucidated. There are still questions about the exact nature of the nucleotide hydrolysis cycle and how it regulates the interaction of Get3 with the sorting complex, TA proteins and the membrane receptor. Models for the molecular interactions of Get3 with TA protein await more structural and *in vivo* studies. The role of the GET pathway and TA protein translocation outside of eukaryotic organisms is a new frontier for study. Future studies of the GET pathway will serve to increase our understanding of the fundamental process of protein translocation.

BIBLIOGRAPHY

- Abell, B.M., Pool, M.R., Schlenker, O., Sinning, I., and High, S. (2004). Signal recognition particle mediates post-translational targeting in eukaryotes. *EMBO J.* *23*, 2755–2764.
- Adams, P.D., Grosse-Kunstleve, R.W., Hung, L.-W., Ioerger, T.R., McCoy, A.J., Moriarty, N.W., Read, R.J., Sacchettini, J.C., Sauter, N.K., and Terwilliger, T.C. (2002). PHENIX: building new software for automated crystallographic structure determination. *Acta Crystallogr. D Biol. Crystallogr.* *58*, 1948–1954.
- Ahmadian, M.R., Stege, P., Scheffzek, K., and Wittinghofer, A. (1997). Confirmation of the arginine-finger hypothesis for the GAP-stimulated GTP-hydrolysis reaction of Ras. *Nat. Struct. Biol.* *4*, 686–689.
- Andreeva, A., Howorth, D., Chandonia, J.-M., Brenner, S.E., Hubbard, T.J.P., Chothia, C., and Murzin, A.G. (2008). Data growth and its impact on the SCOP database: new developments. *Nucleic Acids Res.* *36*, D419–D425.
- Auld, K., Hitchcock, A., Doherty, H., and Frietze, S. (2006). The conserved ATPase Get3/Arr4 modulates the activity of membrane-associated proteins in *Saccharomyces cerevisiae*. *Genetics* *174*, 215–227.
- Battle, A., Jonikas, M.C., Walter, P., Weissman, J.S., and Koller, D. (2010). Automated identification of pathways from quantitative genetic interaction data. *Mol. Syst. Biol.* *6*, 379.
- Battye, T.G.G., Kontogiannis, L., Johnson, O., Powell, H.R., and Leslie, A.G.W. (2011). iMOSFLM: a new graphical interface for diffraction-image processing with MOSFLM. *Acta Crystallogr. D Biol. Crystallogr.* *67*, 271–281.

- Beilharz, T., Egan, B., Silver, P.A., Hofmann, K., and Lithgow, T. (2003). Bipartite signals mediate subcellular targeting of tail-anchored membrane proteins in *Saccharomyces cerevisiae*. *J. Biol. Chem.* *278*, 8219–8223.
- Bergfors, T. (2003). Seeds to crystals. *J. Struct. Biol.* *142*, 66–76.
- Bernstein, H.D., Poritz, M.A., Strub, K., Hoben, P.J., Brenner, S., and Walter, P. (1989). Model for signal sequence recognition from amino-acid sequence of 54K subunit of signal recognition particle. *Nature* *340*, 482–486.
- Borgese, N. (2003). The tale of tail-anchored proteins: coming from the cytosol and looking for a membrane. *J. Cell Biol.* *161*, 1013–1019.
- Borgese, N., and Righi, M. (2010). Remote Origins of Tail-Anchored Proteins. *Traffic* *11*, 877–885.
- Borgese, N., Brambillasca, S., and Colombo, S. (2007). How tails guide tail-anchored proteins to their destinations. *Curr. Opin. Cell Biol.* *19*, 368–375.
- Boskovic, J., Soler-Mira, A., García-Cantalejo, J., Ballesta, J., Jiménez, A., and Remacha, M. (1996). The sequence of a 16 691 bp segment of *Saccharomyces cerevisiae* chromosome IV identifies the DUN1, PMT1, PMT5, SRP14 and DPR1 genes, and five new open reading frames. *Yeast* *12*, 1377–1384.
- Bourne, H.R., Sanders, D.A., and McCormick, F. (1991). The GTPase superfamily: conserved structure and molecular mechanism. *Nature* *349*, 117–127.
- Bozkurt, G., Stjepanovic, G., Vilardi, F., Amlacher, S., Wild, K., Bange, G., Favalaro, V., Rippe, K., Hurt, E., Dobberstein, B., et al. (2009). Structural insights into tail-anchored

protein binding and membrane insertion by Get3. *Proc. Natl. Acad. Sci. USA* *106*, 21131–21136.

Bozkurt, G., Wild, K., Amlacher, S., Hurt, E., Dobberstein, B., and Sinning, I. (2010). The structure of Get4 reveals an alpha-solenoid fold adapted for multiple interactions in tail-anchored protein biogenesis. *FEBS Lett.* *584*, 1509–1514.

Brambillasca, S., Yabal, M., Makarow, M., and Borgese, N. (2006). Unassisted translocation of large polypeptide domains across phospholipid bilayers. *J. Cell Biol.* *175*, 767–777.

Chang, Y.-W., Chuang, Y.-C., Ho, Y.-C., Cheng, M.-Y., Sun, Y.-J., Hsiao, C.-D., and Wang, C. (2010). Crystal structure of Get4/Get5 complex and its interactions with Sgt2, Get3 and Ydj1. *J. Biol. Chem.* *285*, 9962–9970.

Chang, Y.-W., Lin, T.-W., Li, Y.-C., Huang, Y.-S., Sun, Y.-J., and Hsiao, C.-D. (2012). Interaction Surface and Topology of Get3/Get4/Get5, a Complex Involved in Targeting Tail-Anchored Proteins to the Endoplasmic Reticulum. *J. Biol. Chem.* *287*, 4783–4789.

Chartron, J., Suloway, C.J.M., Zaslaver, M., and Clemons, W.M. (2010). Structural characterization of the Get4/Get5 complex and its interaction with Get3. *Proc. Natl. Acad. Sci. USA* *107*, 12127–12132.

Chartron, J.W., Gonzalez, G.M., and Clemons, W.M. (2011). A structural model of the Sgt2 protein and its interactions with chaperones and the Get4/Get5 complex. *J. Biol. Chem.* *286*, 34325–34334.

Chartron, J.W., Vandervelde, D.G., Rao, M., and Clemons, W.M. (2012). The Get5 carboxyl terminal domain is a novel dimerization motif that tethers an extended Get4/Get5 complex. *J. Biol. Chem.* *In press*.

Ching, M.H., Kaur, P., Karkaria, C.E., Steiner, R.F., and Rosen, B.P. (1991). Substrate-induced dimerization of the ArsA protein, the catalytic component of an anion-translocating ATPase. *J. Biol. Chem.* *266*, 2327–2332.

Collaborative Computational Project Number 4. (1994). The CCP4 Suite: Programs for Protein Crystallography. *Acta Crystallogr. D Biol. Crystallogr.* *50*, 760–763.

Copic, A., Dorrington, M., Pagant, S., Barry, J., Lee, M.C.S., Singh, I., Hartman, J.L., and Miller, E.A. (2009). Genomewide analysis reveals novel pathways affecting endoplasmic reticulum homeostasis, protein modification and quality control. *Genetics* *182*, 757–769.

Cordell, S.C., and Löwe, J. (2001). Crystal structure of the bacterial cell division regulator MinD. *FEBS Lett.* *492*, 160–165.

Costanzo, M., Baryshnikova, A., Bellay, J., Kim, Y., Spear, E.D., Sevier, C.S., Ding, H., Koh, J.L.Y., Toufighi, K., Mostafavi, S., et al. (2010). The Genetic Landscape of a Cell. *Science* *327*, 425–431.

Dailey, H.A., and Strittmatter, P. (1978). Structural and functional properties of the membrane binding segment of cytochrome b5. *J. Biol. Chem.* *253*, 8203–8209.

de Boer, P.A., Crossley, R.E., Hand, A.R., and Rothfield, L.I. (1991). The MinD protein is a membrane ATPase required for the correct placement of the *Escherichia coli* division site. *EMBO J.* *10*, 4371–4380.

Delano, W.L. The PyMOL Molecular Graphics System. Schrödinger, LLC.

Derewenda, Z.S., and Vekilov, P.G. (2006). Entropy and surface engineering in protein crystallization. *Acta Crystallogr. D Biol. Crystallogr.* *62*, 116–124.

Doudna, J.A., and Batey, R.T. (2004). Structural insights into the signal recognition particle. *Annu. Rev. Biochem.* *73*, 539–557.

Dunham, T.D., Xu, W., Funnell, B.E., and Schumacher, M.A. (2009). Structural basis for ADP-mediated transcriptional regulation by P1 and P7 ParA. *EMBO J.* *28*, 1792–1802.

Ebersbach, G., and Gerdes, K. (2005). Plasmid segregation mechanisms. *Annu. Rev. Genet.* *39*, 453–479.

Emsley, P., and Cowtan, K. (2004). Coot: model-building tools for molecular graphics. *Acta Crystallogr. D Biol. Crystallogr.* *60*, 2126–2132.

Evans, P. (2006). Scaling and assessment of data quality. *Acta Crystallogr. D Biol. Crystallogr.* *62*, 72–82.

Favaloro, V., Spasic, M., Schwappach, B., and Dobberstein, B. (2008). Distinct targeting pathways for the membrane insertion of tail-anchored (TA) proteins. *J. Cell Sci.* *121*, 1832–1840.

Favaloro, V., Vilaridi, F., Schlecht, R., Mayer, M.P., and Dobberstein, B. (2010). Asna1/TRC40-mediated membrane insertion of tail-anchored proteins. *J. Cell Sci.* *123*, 1522–1530.

Freymann, D.M., Keenan, R.J., Stroud, R.M., and Walter, P. (1997). Structure of the conserved GTPase domain of the signal recognition particle. *Nature* *385*, 361–364.

- Georgiadis, M.M., Komiya, H., Chakrabarti, P., Woo, D., Kornuc, J.J., and Rees, D.C. (1992). Crystallographic structure of the nitrogenase iron protein from *Azotobacter vinelandii*. *Science* 257, 1653–1659.
- Ghosh, S.K., Hajra, S., Paek, A., and Jayaram, M. (2006). Mechanisms for chromosome and plasmid segregation. *Annu. Rev. Biochem.* 75, 211–241.
- Hann, B.C., and Walter, P. (1991). The signal recognition particle in *S. cerevisiae*. *Cell* 67, 131–144.
- Hayashi, I., Oyama, T., and Morikawa, K. (2001). Structural and functional studies of MinD ATPase: implications for the molecular recognition of the bacterial cell division apparatus. *EMBO J.* 20, 1819–1828.
- Hegde, R.S., and Keenan, R.J. (2011). Tail-anchored membrane protein insertion into the endoplasmic reticulum. *Nat. Rev. Mol. Cell. Biol.* 12, 787–798.
- Hemmingsson, O., Zhang, Y., Still, M., and Naredi, P. (2008). ASNA1, an ATPase targeting tail-anchored proteins, regulates melanoma cell growth and sensitivity to cisplatin and arsenite. *Cancer Chemother. Pharmacol.* 63, 491–499.
- Heras, B., and Martin, J.L. (2005). Post-crystallization treatments for improving diffraction quality of protein crystals. *Acta Crystallogr. D Biol. Crystallogr.* 61, 1173–1180.
- Hoover, D.M., and Lubkowski, J. (2002). DNAWorks: an automated method for designing oligonucleotides for PCR-based gene synthesis. *Nucleic Acids Res.* 30, e43.
- Howard, J.B., and Rees, D.C. (1994). Nitrogenase: a nucleotide-dependent molecular switch. *Annu. Rev. Biochem.* 63, 235–264.

- Howard, J.B., Davis, R., Moldenhauer, B., Cash, V.L., and Dean, D. (1989). Fe:S cluster ligands are the only cysteines required for nitrogenase Fe-protein activities. *J. Biol. Chem.* *264*, 11270–11274.
- Hu, J., Li, J., Qian, X., Denic, V., and Sha, B. (2009). The crystal structures of yeast Get3 suggest a mechanism for tail-anchored protein membrane insertion. *PLoS ONE* *4*, e8061.
- Hu, Z., and Lutkenhaus, J. (2003). A conserved sequence at the C-terminus of MinD is required for binding to the membrane and targeting MinC to the septum. *Mol. Microbiol.* *47*, 345–355.
- Jonikas, M.C., Collins, S.R., Denic, V., Oh, E., Quan, E.M., Schmid, V., Weibezahn, J., Schwappach, B., Walter, P., Weissman, J.S., et al. (2009a). Comprehensive Characterization of Genes Required for Protein Folding in the Endoplasmic Reticulum. *Science* *323*, 1693–1697.
- Kabsch, W. (2010). XDS. *Acta Crystallogr. D Biol. Crystallogr.* *66*, 125–132.
- Kalbfleisch, T., Cambon, A., and Wattenberg, B.W. (2007). A bioinformatics approach to identifying tail-anchored proteins in the human genome. *Traffic* *8*, 1687–1694.
- Kao, G., Nordenson, C., Still, M., Rönnlund, A., Tuck, S., and Naredi, P. (2007). ASNA-1 positively regulates insulin secretion in *C. elegans* and mammalian cells. *Cell* *128*, 577–587.
- Kaur, P., and Rosen, B.P. (1992). Mutagenesis of the C-terminal nucleotide-binding site of an anion-translocating ATPase. *J. Biol. Chem.* *267*, 19272–19277.
- Keenan, R.J., Freymann, D.M., Walter, P., and Stroud, R.M. (1998). Crystal structure of the signal sequence binding subunit of the signal recognition particle. *Cell* *94*, 181–191.

Keenan, R.J., Freymann, D.M., Stroud, R.M., and Walter, P. (2001). The signal recognition particle. *Annu. Rev. Biochem.* 70, 755–775.

Kinch, L.N., Saier, M.H., Jr, and Grishin, N.V. (2002). Sec61 β – a component of the archaeal protein secretory system. *Trends Biochem. Sci.* 27, 170–171.

Konarev, P.V., Volkov, V.V., Sokolova, A.V., Koch, M.H.J., and Svergun, D.I. (2003). PRIMUS: a Windows PC-based system for small-angle scattering data analysis. *J. Appl. Crystallogr.* 36, 1277–1282.

Koonin, E.V. (1993). A superfamily of ATPases with diverse functions containing either classical or deviant ATP-binding motif. *J. Mol. Biol.* 229, 1165–1174.

Kriechbaumer, V., Shaw, R., Mukherjee, J., Bowsher, C.G., Harrison, A.-M., and Abell, B.M. (2009). Subcellular distribution of tail-anchored proteins in *Arabidopsis*. *Traffic* 10, 1753–1764.

Krissinel, E., and Henrick, K. (2007). Inference of macromolecular assemblies from crystalline state. *J. Mol. Biol.* 372, 774–797.

Krogh, A., Larsson, B., Heijne, von, G., and Sonnhammer, E.L. (2001). Predicting transmembrane protein topology with a hidden Markov model: application to complete genomes. *J. Mol. Biol.* 305, 567–580.

Kurdi-Haidar, B., Aebi, S., Heath, D., Enns, R.E., Naredi, P., Hom, D.K., and Howell, S.B. (1996). Isolation of the ATP-binding human homolog of the *arsA* component of the bacterial arsenite transporter. *Genomics* 36, 486–491.

- Kutay, U., Ahnert-Hilger, G., Hartmann, E., Wiedenmann, B., and Rapoport, T.A. (1995). Transport route for synaptobrevin via a novel pathway of insertion into the endoplasmic reticulum membrane. *EMBO J.* *14*, 217–223.
- Kutay, U., Hartmann, E., and Rapoport, T.A. (1993). A class of membrane proteins with a C-terminal anchor. *Trends Cell. Biol.* *3*, 72–75.
- Larkin, M.A., Blackshields, G., Brown, N.P., Chenna, P., McGettingan, P.A., McWilliam, H., Valentin, F., Wallace, I.M., Wilm, A., Lopez, R., Thompson, J.D., Gibson, T.J., and Higgins, D.G. (2007) Clustal W and Clustal X version 2.0. *Bioinformatics* *21*, 2947–2948.
- Lee, M.J., and Dohlman, H.G. (2008). Coactivation of G protein signaling by cell-surface receptors and an intracellular exchange factor. *Curr. Biol.* *18*, 211–215.
- Leipe, D., Wolf, Y., Koonin, E., and Aravind, L. (2002). Classification and evolution of P-loop GTPases and related ATPases. *J. Mol. Biol.* *317*, 41–72.
- Leonard, T.A., Butler, P.J., and Löwe, J. (2005). Bacterial chromosome segregation: structure and DNA binding of the Soj dimer--a conserved biological switch. *EMBO J.* *24*, 270–282.
- Leslie, A. (1992) Recent changes to the MOSFLM package for processing film and image plate data. *Joint CCP4 and ESF-EAMCB Newsletter on Protein Crystallography* *26*.
- Leznicki, P., Clancy, A., and Schwappach, B. (2010). Bat3 promotes the membrane integration of tail-anchored proteins. *J. Cell Sci.* *123*, 2170–2178.
- Leznicki, P., Warwicker, J., and High, S. (2011). A biochemical analysis of the constraints of tail-anchored protein biogenesis. *Biochem. J.* *436*, 719–727.

- Li, J., and Rosen, B.P. (2000). The linker peptide of the ArsA ATPase. *Mol. Microbiol.* *35*, 361–367.
- Liou, S.-T., and Wang, C. (2005). Small glutamine-rich tetratricopeptide repeat-containing protein is composed of three structural units with distinct functions. *Arch. Biochem. Biophys.* *435*, 253–263.
- Lutkenhaus, J. (2007). Assembly dynamics of the bacterial MinCDE system and spatial regulation of the Z ring. *Annu. Rev. Biochem.* *76*, 539–562.
- Lutkenhaus, J., and Sundaramoorthy, M. (2003). MinD and role of the deviant Walker A motif, dimerization and membrane binding in oscillation. *Mol Microbiol* *48*, 295–303.
- Mandon, E.C., and Gilmore, R. (2007). The tail end of membrane insertion. *Cell* *128*, 1031–1032.
- Mariappan, M., Li, X., Stefanovic, S., Sharma, A., Mateja, A., Keenan, R.J., and Hegde, R.S. (2010). A ribosome-associating factor chaperones tail-anchored membrane proteins. *Nature* *466*, 1120–1124.
- Mariappan, M., Mateja, A., Dobosz, M., Bove, E., Hegde, R.S., and Keenan, R.J. (2011). The mechanism of membrane-associated steps in tail-anchored protein insertion. *Nature* *477*, 61–66.
- Mateja, A., Szlachcic, A., Downing, M.E., Dobosz, M., Mariappan, M., Hegde, R.S., and Keenan, R.J. (2009). The structural basis of tail-anchored membrane protein recognition by Get3. *Nature* *461*, 361–366.

McCoy, A.J. (2007). Solving structures of protein complexes by molecular replacement with Phaser. *Acta Crystallogr. D Biol. Crystallogr.* *63*, 32–41.

McPhillips, T.M., McPhillips, S.E., Chiu, H.-J., Cohen, A.E., Deacon, A.M., Ellis, P.J., Garman, E., Gonzalez, A., Sauter, N.K., Phizackerley, R.P., et al. (2002). Blu-Ice and the Distributed Control System: software for data acquisition and instrument control at macromolecular crystallography beamlines. *J. Synchrotron Radiat.* *9*, 401–406.

Metz, J., Wächter, A., Schmidt, B., Bujnicki, J.M., and Schwappach, B. (2006). The yeast Arr4p ATPase binds the chloride transporter Gef1p when copper is available in the cytosol. *J. Biol. Chem.* *281*, 410–417.

Milburn, M.V., Tong, L., deVos, A.M., Brünger, A., Yamaizumi, Z., Nishimura, S., and Kim, S.H. (1990). Molecular switch for signal transduction: structural differences between active and inactive forms of protooncogenic ras proteins. *Science* *247*, 939–945.

Montoya, G., Svensson, C., Lührink, J., and Sinning, I. (1997). Crystal structure of the NG domain from the signal-recognition particle receptor FtsY. *Nature* *385*, 365–368.

Mukhopadhyay, R., Ho, Y.-S., Swiatek, P.J., Rosen, B.P., and Bhattacharjee, H. (2006). Targeted disruption of the mouse Asna1 gene results in embryonic lethality. *FEBS Lett.* *580*, 3889–3894.

Netz, D.J.A., Pierik, A.J., Stuempfig, M., Muehlenhoff, U., and Lill, R. (2007). The Cfd1-Nbp35 complex acts as a scaffold for iron-sulfur protein assembly in the yeast cytosol. *Nat. Chem. Biol.* *3*, 278–286.

- Ogura, T., and Wilkinson, A.J. (2001). AAA+ superfamily ATPases: common structure--diverse function. *Genes Cells* 6, 575–597.
- Painter, J., and Merritt, E. (2006a). Optimal description of a protein structure in terms of multiple groups undergoing TLS motion. *Acta Crystallogr. D Biol. Crystallogr.* 62, 439–450.
- Painter, J., and Merritt, E.A. (2006b). TLSMDweb server for the generation of multi-group TLS models. *J. Appl. Crystallogr.* 39, 109–111.
- Pedrazzini, E. (2009). Tail-Anchored Proteins in Plants. *J. Plant Biol.* 52, 88–101.
- Pettersen, E., Goddard, T., Huang, C., Couch, G., Greenblatt, D., Meng, E., and Ferrin, T. (2004). UCSF chimera - A visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–1612.
- Pratto, F., Cicek, A., Weihofen, W.A., Lurz, R., Saenger, W., and Alonso, J.C. (2008). *Streptococcus pyogenes* pSM19035 requires dynamic assembly of ATP-bound ParA and ParB on parS DNA during plasmid segregation. *Nucleic Acids Res.* 36, 3676–3689.
- Putnam, C.D., Hammel, M., Hura, G.L., and Tainer, J.A. (2007). X-ray solution scattering (SAXS) combined with crystallography and computation: defining accurate macromolecular structures, conformations and assemblies in solution. *Q. Rev. Biophys.* 40, 191–285.
- Rabu, C., Schmid, V., and Schwappach, B. (2009). Biogenesis of tail-anchored proteins: the beginning for the end? *J. Cell Sci.* 122, 3605–3612.
- Sakai, N., Yao, M., Itou, H., Watanabe, N., Yumoto, F., Tanokura, M., and Tanaka, I. (2001). The three-dimensional structure of septum site- determining protein MinD from *Pyrococcus horikoshii* OT3 in complex with Mg-ADP. *Structure* 9, 817–826.

Sanner, M., Olson, A., and Spehner, J. (1996). Reduced surface: An efficient way to compute molecular surfaces. *Biopolymers* *38*, 305–320.

Santangelo, T.J., Čuboňová, L., James, C.L., and Reeve, J.N. (2007). TFB1 or TFB2 is sufficient for *Thermococcus kodakaraensis* viability and for basal transcription in vitro. *J. Mol. Biol.* *367*, 344–357.

Saraste, M., Sibbald, P.R., and Wittinghofer, A. (1990). The P-loop--a common motif in ATP- and GTP-binding proteins. *Trends Biochem. Sci.* *15*, 430–434.

Schaffitzel, C., Oswald, M., Berger, I., Ishikawa, T., Abrahams, J.P., Koerten, H.K., Koning, R.I., and Ban, N. (2006). Structure of the *E. coli* signal recognition particle bound to a translating ribosome. *Nature* *444*, 503–506.

Scheffzek, K., Ahmadian, M., Kabsch, W., Wiesmuller, L., Lautwein, A., Schmitz, F., and Wittinghofer, A. (1997). The Ras-RasGAP complex: Structural basis for GTPase activation and its loss in oncogenic Ras mutants. *Science* *277*, 333–338.

Schindelin, H., Kisker, C., Schlessman, J.L., Howard, J.B., and Rees, D.C. (1997). Structure of ADP·AlF₄⁻-stabilized nitrogenase complex and its implications for signal transduction. *Nature* *387*, 370–376.

Schlessman, J.L., Woo, D., Joshua-Tor, L., Howard, J.B., and Rees, D.C. (1998). Conformational variability in structures of the nitrogenase iron proteins from *Azotobacter vinelandii* and *Clostridium pasteurianum*. *J. Mol. Biol.* *280*, 669–685.

Schuldiner, M., Collins, S.R., Thompson, N.J., Denic, V., Bhamidipati, A., Punna, T., Ihmels, J., Andrews, B., Boone, C., Greenblatt, J.F., et al. (2005). Exploration of the function and

organization of the yeast early secretory pathway through an epistatic miniarray profile. *Cell* *123*, 507–519.

Schuldiner, M., Metz, J., Schmid, V., Denic, V., Rakwalska, M., Schmitt, H.D., Schwappach, B., and Weissman, J.S. (2008). The GET complex mediates insertion of tail-anchored proteins into the ER membrane. *Cell* *134*, 634–645.

Shan, S.-O., and Walter, P. (2005). Co-translational protein targeting by the signal recognition particle. *FEBS Lett* *579*, 921–926.

Sheldrick, G.M. (2008). A short history of SHELX. *Acta Crystallogr. A* *64*, 112–122.

Shen, J., Hsu, C.-M., Kang, B.-K., Rosen, B.P., and Bhattacharjee, H. (2003). The *Saccharomyces cerevisiae* Arr4p is involved in metal and heat tolerance. *Biometals* *16*, 369–378.

Sherrill, J., Mariappan, M., Dominik, P., Hegde, R.S., and Keenan, R.J. (2011). A conserved archaeal pathway for tail-anchored membrane protein insertion. *Traffic* *12*, 1119–1123.

Simpson, P.J., Schwappach, B., Dohlman, H.G., and Isaacson, R.L. (2010). Structures of Get3, Get4, and Get5 provide new models for TA membrane protein targeting. *Structure* *18*, 897–902.

Sprang, S.R. (1997). G protein mechanisms: insights from structural analysis. *Annu Rev Biochem* *66*, 639–678.

Steel, G.J., Brownsword, J., and Stirling, C.J. (2002). Tail-anchored protein insertion into yeast ER requires a novel posttranslational mechanism which is independent of the SEC machinery. *Biochemistry* *41*, 11914–11920.

- Stefanovic, S., and Hegde, R.S. (2007). Identification of a targeting factor for posttranslational membrane protein insertion into the ER. *Cell* *128*, 1147–1159.
- Stefer, S., Reitz, S., Wang, F., Wild, K., Pang, Y.-Y., Schwarz, D., Bomke, J., Hein, C., Löhr, F., Bernhard, F., et al. (2011). Structural Basis for Tail-Anchored Membrane Protein Biogenesis by the Get3-Receptor Complex. *Science* *333*, 758–762.
- Strong, M., Sawaya, M.R., Wang, S., Phillips, M., Cascio, D., and Eisenberg, D. (2006). Toward the structural genomics of complexes: crystal structure of a PE/PPE protein complex from *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. USA* *103*, 8060–8065.
- Studier, F.W. (2005). Protein production by auto-induction in high-density shaking cultures. *Protein Expr. Purif.* *41*, 207–234.
- Suloway, C.J., Rome, M.E., and Clemons, W.M. (2012). Tail-anchor targeting by a Get3 tetramer: the structure of an archaeal homologue. *EMBO J.* *31*, 707–719.
- Suloway, C.J.M., Chartron, J.W., Zaslaver, M., and Clemons, W.M. (2009). Model for eukaryotic tail-anchored protein binding based on the structure of Get3. *Proc. Natl. Acad. Sci. U.S.A.* *106*, 14849–14854.
- Sulston, J., Du, Z., Thomas, K., Wilson, R., Hillier, L., Staden, R., Halloran, N., Green, P., Thierry-mieg, J., Qiu, L., et al. (1992). The *C. elegans* genome sequencing project: a beginning. *Nature* *356*, 37–41.
- Svergun, D.I. (1992). Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *J. Appl. Crystallogr.* *25*, 495–503.

- Svergun, D.I. (1999). Restoring Low Resolution Structure of Biological Macromolecules from Solution Scattering Using Simulated Annealing. *Biophys. J.* **76**, 2879–2886.
- Takagaki, Y., Radhakrishnan, R., Wirtz, K., and Khorana, H.G. (1983). The membrane-embedded segment of cytochrome b5 as studied by cross-linking with photoactivatable phospholipids. II. The nontransferable form. *J. Biol. Chem.* **258**, 9136.
- Terwilliger, T.C., and Berendzen, J. (1999). Automated MAD and MIR structure solution. *Acta Crystallogr. D Biol. Crystallogr.* **55**, 849–861.
- Tezcan, F.A., Kaiser, J.T., Mustafi, D., Walton, M.Y., Howard, J.B., and Rees, D.C. (2005). Nitrogenase complexes: multiple docking sites for a nucleotide switch protein. *Science* **309**, 1377–1380.
- Tobaben, S., Varoqueaux, F., Brose, N., Stahl, B., and Meyer, G. (2003). A brain-specific isoform of small glutamine-rich tetratricopeptide repeat-containing protein binds to Hsc70 and the cysteine string protein. *J. Biol. Chem.* **278**, 38376–38383.
- Tseng, Y.-Y., Yu, C.-W., and Liao, V.H.-C. (2007). *Caenorhabditis elegans* expresses a functional ArsA. *FEBS J.* **274**, 2566–2572.
- Van Den Berg, B., Clemons, W.M., Collinson, I., Modis, Y., Hartmann, E., Harrison, S.C., and Rapoport, T.A. (2004). X-ray structure of a protein-conducting channel. *Nature* **427**, 36–44.
- Vilardi, F., Lorenz, H., and Dobberstein, B. (2011). WRB is the receptor for TRC40/Asna1-mediated insertion of tail-anchored proteins into the ER membrane. *J. Cell Sci.* **124**, 1301–1307.

Walker, J.E., Saraste, M., Runswick, M.J., and Gay, N.J. (1982). Distantly related sequences in the alpha- and beta-subunits of ATP synthase, myosin, kinases and other ATP-requiring enzymes and a common nucleotide binding fold. *EMBO J.* *1*, 945–951.

Walmsley, A.R., Zhou, T., Borges-Walmsley, M.I., and Rosen, B.P. (1999). The ATPase Mechanism of ArsA, the Catalytic Subunit of the Arsenite Pump. *J. Biol. Chem.* *274*, 16153–16161.

Wang, F., Brown, E.C., Mak, G., Zhuang, J., and Denic, V. (2010). A chaperone cascade sorts proteins for posttranslational membrane insertion into the endoplasmic reticulum. *Mol. Cell.* *40*, 159–171.

Wang, H.W., Lu, Y.J., Li, L.J., Liu, S., Wang, D.N., and Sui, S. (2000). Trimeric ring-like structure of ArsA ATPase. *FEBS Lett.* *469*, 105–110.

Wu, C., Amrani, N., Jacobson, A., and Sachs, M.S. (2007). The use of fungal in vitro systems for studying translational regulation. *Method Enzymol.* *429*, 203–225.

Wu, C.C., and Yates, J.R. (2003). The application of mass spectrometry to membrane proteomics. *Nat. Biotechnol.* *21*, 262–267.

Yabal, M., Brambillasca, S., Soffientini, P., Pedrazzini, E., Borgese, N., and Makarow, M. (2003). Translocation of the C terminus of a tail-anchored protein across the endoplasmic reticulum membrane in yeast mutants defective in signal peptide-driven translocation. *J. Biol. Chem.* *278*, 3489–3496.

- Yamagata, A., Mimura, H., Sato, Y., Yamashita, M., Yoshikawa, A., and Fukai, S. (2010). Structural insight into the membrane insertion of tail-anchored proteins by Get3. *Genes Cells* 15, 29–41.
- Zhou, T., Radaev, S., Rosen, B.P., and Gatti, D.L. (2000). Structure of the ArsA ATPase: the catalytic subunit of a heavy metal resistance pump. *EMBO J.* 19, 4838–4845.
- Zhou, T., Radaev, S., Rosen, B.P., and Gatti, D.L. (2001). Conformational changes in four regions of the Escherichia coli ArsA ATPase link ATP hydrolysis to ion translocation. *J. Biol. Chem.* 276, 30414–30422.
- Zúñiga, S., Boskovic, J., Jiménez, A., Ballesta, J.P., and Remacha, M. (1999). Disruption of six *Saccharomyces cerevisiae* novel genes and phenotypic analysis of the deletants. *Yeast* 15, 945–953.

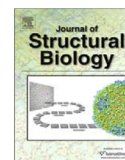
APPENDIX A

Journal of Structural Biology 167 (2009) 11–18



Contents lists available at ScienceDirect

Journal of Structural Biology

journal homepage: www.elsevier.com/locate/yjsbiFully automated, sequential tilt-series acquisition with *Leginon*Christian Suloway^a, Jian Shi^a, Anchi Cheng^c, James Pulokas^c, Bridget Carragher^c, Clinton S. Potter^c, Shawn Q. Zheng^{d,e}, David A. Agard^{d,e}, Grant J. Jensen^{a,b,*}^a Division of Biology, California Institute of Technology, 1200 E. California Blvd., Pasadena, CA 91125, USA^b Howard Hughes Medical Institute, California Institute of Technology, 1200 E. California Blvd., Pasadena, CA 91125, USA^c The National Resource for Automated Molecular Microscopy, Department of Cell Biology, The Scripps Research Institute, 10550 North Torrey Pines Rd., La Jolla, CA 92037, USA^d W.M. Keck Advanced Microscopy Laboratory, University of California, San Francisco, CA 94158-2517, USA^e Howard Hughes Medical Institute, University of California, San Francisco, CA 94158-2517, USA

ARTICLE INFO

Article history:

Received 21 October 2008

Received in revised form 27 March 2009

Accepted 31 March 2009

Available online 8 April 2009

Keywords:

TEM

Cryo-electron microscopy

Electron microscopy

Automation

Tomography

ABSTRACT

Electron tomography has become a uniquely powerful tool for investigating the structures of individual cells, viruses, and macromolecules. Data collection is, however, time consuming and requires expensive instruments. To optimize productivity, we have incorporated one of the existing tilt-series acquisition programs, *UCSF Tomo*, into the well-developed automatic electron microscopy data collection package *Leginon* to enable fully automatic, sequential tilt-series acquisition. Here we describe how *UCSF Tomo* was integrated into *Leginon*, what users must do to set up a data collection session, how the automatic collection proceeds, how archived data about the process can be accessed and used, and how the software has been tested.

© 2009 Elsevier Inc. All rights reserved.

1. Introduction

Electron tomography (ET) is the highest resolution technique available today for observing the 3D structure of unique biological objects such as cells, many viruses, and certain flexible macromolecules, and is also being used increasingly in various “Materials science” applications (Jensen and Briegel, 2007; Lučić et al., 2005; McIntosh et al., 2005). In ET, a series of projection images is recorded through a specimen with a transmission electron microscope as the specimen is incrementally tilted around one and then sometimes a second axis. 3D reconstructions, or “tomograms,” are then calculated from such “tilt-series” through back-projection or other reconstruction algorithms. Specimens can be imaged at either ambient or cryogenic temperatures, enabling high-resolution visualization of large volumes of fixed samples (through serial section montaging [Marsh, 2005]) as well as more detailed analysis of smaller samples preserved in a near-native, frozen-hydrated state (Murphy and Jensen, 2005).

While the basic idea of ET has been around for decades (DeRosier and Moore, 1970; Hart, 1968; Hoppe et al., 1968), technical challenges have until recently prevented its widespread use in all but

a few dedicated laboratories. Because no goniometer is mechanically perfect and samples are never exactly at eucentric height, as the sample is tilted it moves both laterally and vertically within the column. Instrumental advances such as CCD cameras and electronic controls were therefore needed before software could be written to automatically tilt the sample, apply beam and image shifts to keep the specimen under the beam and its image centered on the CCD, and adjust focus throughout the tilt-series (Koster et al., 1992). A number of software packages for this are now available, including *SerialEM* (Mastronarde, 2005), *TOM* (Nickell et al., 2005), *Xplore3D* (FEI Company), and *UCSF Tomo* (Zheng et al., 2004), offering a variety of different schemes for data collection and specimen tracking. Tilt-series usually take an hour or two to record, so in a typical working day operators can only acquire a few. Technical problems often ruin tilt-series, requiring additional targets to be found and imaged. Compounding the challenge, many biological questions require the comparison of tens or hundreds of successful tomograms (Briegel et al., 2008; Li et al., 2007). Serial section and montage tomography applications are even more demanding (Marsh, 2005).

Two opportunities for increasing throughput are first to facilitate the process of finding the best targets on a grid, and second, to automate the process of advancing from one target to another during a session. This would free the user from all but the first hour or so of the session when the targets were chosen, and then allow efficient, uninterrupted data collection to proceed through nights

* Corresponding author. Address: Howard Hughes Medical Institute, California Institute of Technology, 1200 E. California Blvd., Pasadena, CA 91125, USA. Fax: +1 626 395 5730.

E-mail address: jensen@caltech.edu (G.J. Jensen).

and weekends without the constraints of user endurance. Several good packages that assist in the process of target selection and sequential imaging (Lei and Frank, 2005; Oostergetel et al., 1998; Potter et al., 1999; Shi et al., 2008; Suloway et al., 2005; Zhang et al., 2001) have already been developed for single particle imaging, and the *SerialEM* and *TOM* tomography packages have already been extended to support sequential tilt-series acquisition. Based on a different idea of how to improve data collection, *UCSF Tomo* was recently enhanced to produce reconstructions in real-time, allowing operators to evaluate results and adjust their collection strategies accordingly mid-session (Zheng et al., 2007). Here we describe our efforts to integrate *UCSF Tomo*'s fast, predictive tracking algorithm into *Leginon*, a well-developed “single-particle” and “2D crystallographic” automatic data collection package with sophisticated targeting, imaging, management, and querying tools. Our choice to incorporate *UCSF Tomo* into *Leginon* is part of a long-term plan to eventually provide a single package (*Leginon*) that will support all the major three-dimensional electron microscopy data collection protocols (2D crystallography, single particle analysis, and now tomography) on all the major microscope brands.

2. Integration of *UCSF Tomo* into *Leginon*

The *Leginon* software package (Suloway et al., 2005) is comprised of “nodes,” defined as individual modules capable of performing tasks such as acquiring an image from the microscope or finding a target within an image. In the code, each node is implemented as a separate class, and so can have inheritance relationships with others. Nodes communicate through “events” that deliver information like a particular target location. The types of

nodes that are active and the order of events that are transmitted between them define the data collection process, which is also known as an “application.” Existing *Leginon* applications include, for instance, fully and partially automated single-particle data collection. These applications proceed by first assembling an “atlas” of tiled, low magnification images; searching the atlas for promising regions/targets; and finally recording higher magnification images of each target one after another. The appropriate magnification, beam intensity, and beam shifts for each step are stored as part of data referred to as “presets,” which are analogous to the “states” commonly used in low dose kits.

In order to incorporate *UCSF Tomo* into *Leginon*, a new application with new graphical user interface elements was created called “*MSI Tomography*,” endowed with all the same atlas-acquisition and sequential target imaging capabilities of previous *Leginon* applications. The key difference is that after a series of targets is identified from a low magnification atlas (Fig. 1), instead of recording a single high magnification projection image of each target (like is done for single particle applications), *MSI Tomography* records an entire tilt-series. To accomplish this a new subclass of the existing “*Acquisition*” node/class was defined named the “*Tomography*” node/class, which manages the collection of a tilt-series. In addition to all the members of its parent *Acquisition* class, the *Tomography* class uses five helper classes which were ported directly from *UCSF Tomo* (Microsoft Visual Studio C++ to Python). The fast performance of the *Tomography* node was maintained through the use of C/C++ Python extensions and the NumPy and SciPy packages for computationally intense calculations. The *Tomography* node's user interface, also adapted from *UCSF Tomo*, contains three image panels to show the last two images recorded and their cross-correlation (Fig. 2).

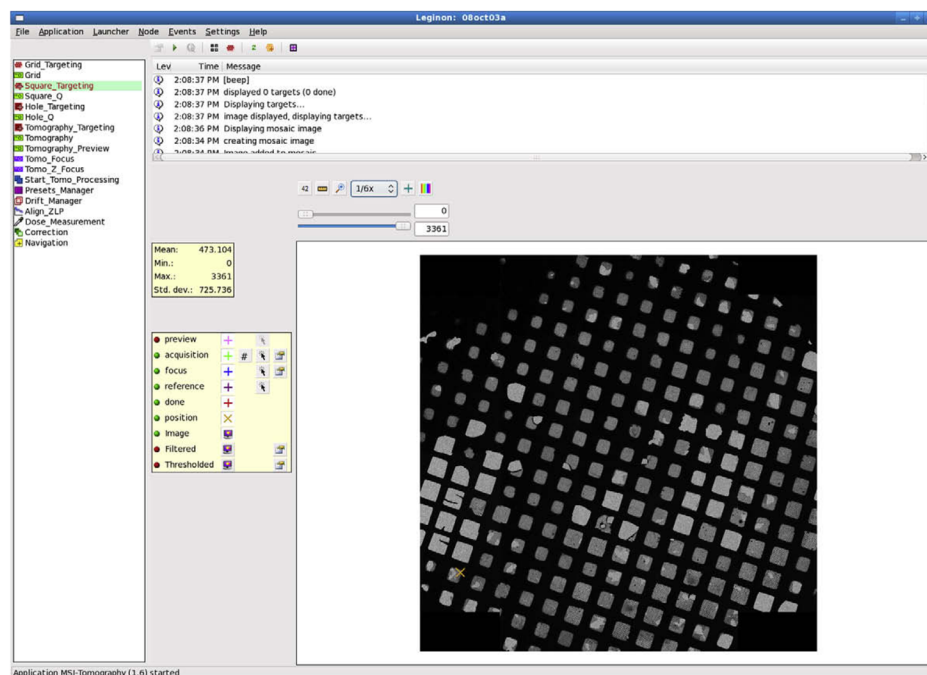


Fig. 1. Screenshot of *MSI Tomography*'s graphical user interface during target selection. General *Leginon* functions are provided along the top menu. The various *Leginon* nodes available in the new *MSI Tomography* application are listed on the left, and a message log records events and alerts in the upper right. Once a specific node is selected (here the *Square Targeting* node), additional node-specific functions appear on a button bar below the menu and appropriate images appear in the main panel to the lower right, surrounded by further buttons and tools. The tools shown are for selecting and further inspecting potential target, focusing, and reference locations.

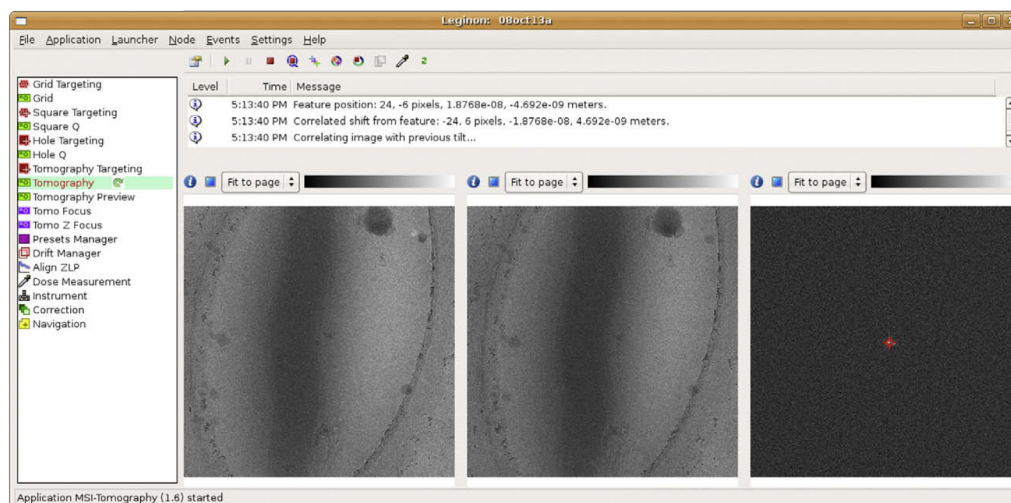


Fig. 2. Screenshot of MSI Tomography's graphical user interface during tilt-series acquisition. The *Tomography* node is selected in the node list (left) to display the newly added user interface. The message log (top center) displays information about the tilt-series' progression, including measurements and predictions. Three image viewers are included as in UCSF Tomo (bottom center): the left image displays the previous tilt image, the center image is the current tilt image, and the right image displays their cross-correlation with a peak identified.

In addition to the new *Tomography* node, two other nodes, *MeasureDose* and *AlignZeroLossPeak*, were added to *Leginon* as well as a new global position known as the "Reference." Assuming a large empty hole in the grid is chosen as the Reference, the *MeasureDose* node can be used to assess long-term fluctuations in beam intensity by acquiring images of the uninterrupted beam. The *AlignZeroLossPeak* node is also meant to be executed at the Reference position, where it adjusts for slit drifts in Gatan Imaging Filters by calling a DigitalMicrograph (Gatan) script of the same name. Finally, several new web-based viewing tools were written to track the application's progress and analyze the images and metadata captured in the *Leginon* database.

3. User tasks

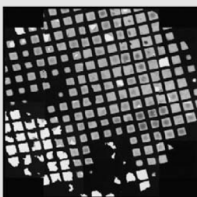
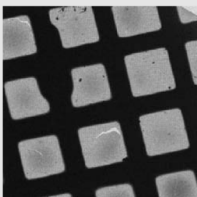
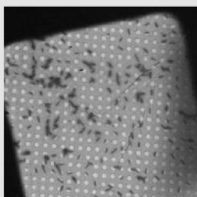
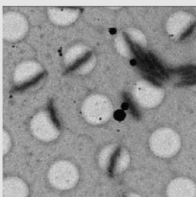
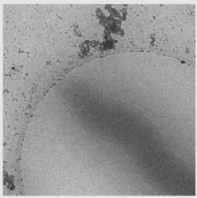
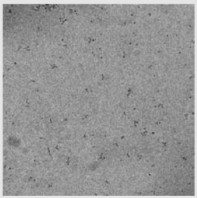
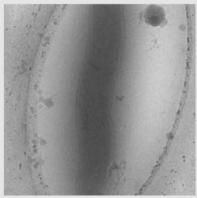
The main goal of this work was to consolidate all the user tasks into a single, short "setup and target selection" process at the beginning of the data collection session. The process of calibration, preset setup, and multi-scale target selection and imaging are similar to other *Leginon* MSI applications, where targets are selected in low magnification images and then imaged sequentially at higher magnifications. The following text is not meant to be a user manual, but is intended to give the reader a sense of the workflow and time requirements.

As for the other *Leginon* applications, the magnitude and direction of the microscope's actual response to various commands like beam and image shifts and stage movements must be calibrated periodically, especially after the instrument (and in particular the goniometer) is serviced (Suloway et al., 2005). These calibrations can take hours to perform, but are then stored and can be used by all *Leginon* applications for weeks. Individual data collection sessions begin by setting up the six or more presets that will be used for finding and later re-centering targets, focusing and making other adjustments, and acquiring the actual tilt-series (Table 1). Generally presets are not created *de novo*, but they are imported from previous experiments and fine-tuned before each session. The electron beam position relative to the specimen and detector changes significantly over time, and therefore requires fine-tuning

each time the microscope is used. The intensity of the electron beam also changes over time, and so generally also requires adjustment, especially in presets where dose and beam size are critical. Establishing and fine-tuning the presets usually takes 30–60 min, depending on how appropriate and well-adjusted the existing presets already are.

Once the presets are established, targets are selected. Whereas without automation, dose-sensitive cryo-EM targets are typically just "wandered into" and chosen based on limited sampling of the grid, automation allows a user to quickly survey large regions of the grid to find and prioritize the very best available targets. First, an atlas of the entire grid (or at least the region that will still be visible at high tilt) is generated by tiling low magnification ("Grid" preset) images into a montage (Table 1, first column, and Fig. 1). A few of the grid squares with the best ice thickness are then chosen by the user, a higher magnification image of each square is recorded with the "Square" preset, and a rough stage-z-height adjustment is performed. Depending on the size, abundance, and nature of the sample, additional images at progressively higher magnification are then used to find and prioritize the best targets. For micron-long bacterial cells, for instance, promising regions of squares are chosen and the "Hole" preset is then used to obtain images at high enough magnification that the boundaries of individual cells can be seen draped across holes in the carbon support. At this point potential targets are marked for further inspection and the "Tomography Preview" preset is used to record an image of each one, generally at the same or slightly lower magnification as the tilt-series will be acquired, but with much higher defocus to enhance contrast and heavily binned to minimize sample exposure. Tomography Preview images can be used to distinguish, for instance, a virus particle from a liposome, or the extent of cell constriction in a dividing bacterial cell, before time is invested into a tilt-series. In this way users choose the final targets that will be imaged. In addition to the center point of each target, a nearby position is also marked for fine stage-z-height adjustment and focusing. The time required to select targets is highly variable, and has in our experience ranged from one to several hours.

Table 1Example presets typically used for ET data collection in *Leginon* (assume a 4k × 4k camera).

Preset name	Atlas ^a	Grid	Square	Hole
Example image				
Magnification	–	100×	480×	3k×
Size (pixels/side)	–	512	512	512
Defocus (μm)	0	0	0	~100
Pixel size after binning (nm)	–	840	170	28
Dose delivered to specimen (e [−] /Å ²)	Negligible	Negligible	Negligible	<0.01
Field of view (μm/side)	~2000	430	90	14
Description/use	Selecting viable grid squares	Building the atlas	Selecting regions with good holes	Selecting potential targets and corresponding focus positions
Preset name	Tomography preview	Focus	Tomography	
Example image				
Magnification	22.5k×	27.5k×	27.5×	
Size (pixels/side)	1024	1024	2048	
Defocus (μm)	~20	–	~10	
Pixel size after binning (nm)	1.9	1.6	0.78	
Dose delivered to specimen (e [−] /Å ²)	<0.25	0 ^b	1–3 ^c	
Field of view (μm/side)	2.0	1.6	1.6	
Description/use	Highly defocused to enhance contrast/final selection and prioritization of targets	Adjusting defocus and stage z-height	Acquiring the tilt-series	

^a Atlas is not technically a preset but actually a montage of Grid images.^b Focus targets are selected to be out of the specimen area.^c Tomography preset dose depends on the tilt-series dose and current tilt angle.

Once targets are selected, but before *Leginon* begins sequential tilt-series acquisition, the two lower magnification presets (Grid and Square) are dropped and the remaining presets are readjusted. This is done because whenever *Leginon* moves from one preset to another, it cycles through the entire set of active presets to reduce hysteresis, and unless the lower magnification presets are removed from the list, such cycling would cause the objective lens to be turned on and off frequently. The “Focus” and “Tomography” presets are usually set to the same (high) magnification and are used for focusing and tilt-series acquisition, respectively. Compared to the “Tomography” preset, the “Focus” preset uses a more condensed beam for improved contrast, but within a smaller area to make sure the target is not exposed prematurely. Thus once the Grid and Square presets are dropped, the objective lens will remain in the “SA” mode for the rest of the session as targets are re-centered and imaged one by one. At this point it can be helpful to let the microscope stabilize for 10–20 min, and then after temperatures and currents equilibrate, the objective aperture can be inserted and centered. This second round of preset setup and stabilization usually takes about another half an hour, but then the user can launch sequential tilt-series acquisition and leave.

4. Automatic data collection

Before the first target is imaged and periodically thereafter (for instance every several hours), *Leginon* adjusts for small fluctuations in the intensity and voltage of the beam. The stage is moved to the Reference position on the grid (usually a large empty hole), and using the Tomography preset the *AlignZeroLossPeak* node re-centers the energy slit. Then the *MeasureDose* node re-checks the intensity of the uninterrupted beam so that the exact dose specified by the user is delivered.

Next *Leginon* moves to the focus position adjacent to the first target and sets the stage-z-height to the gross position estimated previously during target selection. The stage-z-height is now set more precisely by the *Focuser* node, again by measuring image displacement at two different alpha-axis stage tilts, but this time at higher magnification (Hole preset) and with progressively larger stage tilt. Once the stage is at the eucentric height, the *Focuser* node further increases the magnification to the Focus preset and adjusts focus and astigmatism by measuring the image displacements induced by electron beam tilt (Koster et al., 1987). Next the stage is moved to the recorded position of the first target and a

“targeting” image is recorded using the Hole preset. This image is cross-correlated with the original (Hole) image where the center of the target was marked to assess stage drift and/or errors in stage movement. This process can be repeated as necessary until the desired stage position accuracy is reached (Yoshioka et al., 2007). Up to this point, the various targets are reached by stage movement as would be the case for manual targeting. Here and throughout data collection, the measurement and correction for drift is handled by *Leginon*'s *DriftManager* node, which stores cumulative measurements of drift and adds them to target coordinates whenever necessary. Finally, before the tilt-series is begun, the backlash that is often released when the stage is first tilted is released. This is done by tilting the stage slightly (for instance to 5°) and then returning the stage to the untilted position in small (for instance 1°) steps. Another (Hole) image is recorded, and the stage movement is compensated for by beam and image shifts to prevent any further (lateral) mechanical movements of the stage. Despite the many images that have already been recorded of the target, because the dose per image in the Hole preset is typically only .01 electrons/Å², the total dose applied is still essentially negligible.

Now that *Leginon* has re-centered the energy slit, calibrated the dose, fine-tuned the stage-z-height, refocused, removed stigmatism, re-centered the target, and released backlash, the actual tilt-series begins. As with *UCSF Tomo*, the first image is recorded with the stage untilted (using the Tomography preset). The stage is then tilted one increment and a second image is recorded. The first two images are cross-correlated, the trajectory of the specimen in x, y, and z is modeled, the specimen is tilted another increment, the predicted beam and image shifts and focus changes are applied, and the next image of the tilt-series is acquired. This process continues as the specimen is progressively tilted (Fig. 3), aided by the fact that the number of data points available to model the specimen's trajectory increases just as the magnitudes of the specimen movements do too at high tilt angles. Unlike *UCSF Tomo*, the *Tomography* node of *MSI Tomography* has the added advantage of being able to load information from previous tilt-series stored in *Leginon*'s MySQL database (see below) to initiate the model for the specimen's trajectory even before the first tilt. Options are therefore given to initiate the model either with user-defined pre-calibrations or with existing models (like the most recent one recorded at that magnification) stored in the database. In addition, tilt axis angle and offset values that normally are excluded in the model fitting can also be fitted dynamically to assist calibration. After the first half of the tilt-series is recorded, the specimen is returned to the untilted position, re-centered, and the process repeated in the opposite tilt direction. Once the tilt-series is completed, the stage is moved to the focus position adjacent to the next target and the entire process repeats until all the targets have been imaged (Fig. 4).

There are a few differences between the original *UCSF Tomo* and its new implementation within *Leginon*. Perhaps most importantly, because *Leginon* manages sequential tilt-series within a single session and a database is maintained with information from all previous sessions, previous tracking models can be compared and used as “seeds” in future runs. To correct for the occasionally large stage shifts that can occur when a specimen is highly tilted in a side-entry cryostage, *Leginon* re-centers targets using iterative stage movements before recording the second half of the tilt-series. Re-centering the target iteratively with stage movements (rather than image and beam shifts) before tilting was also found to make the stage trajectory more consistent. In order to improve cross-correlation performance, images are binned to 512 × 512 pixels and two additional filters are applied: a low-pass filter with a Gaussian kernel of 1.5 and a Wiener filter with an estimated noise 10 times the mean value of the highest 20% of frequencies. An affine transform is used to “untilt” the images along the measured tilt axis.

Experience with *Leginon* has shown that if the same gain reference image is used to normalize two different “raw” images off the CCD, an unhelpful peak can emerge at the origin of their cross-correlation map due to correlations in the gain reference correction itself. As a standard feature, *Leginon*, therefore, records multiple dark and bright field reference images and constructs multiple similar, but independent gain reference images, called “channels.” Whenever cross-correlation maps are calculated, different channels are used to correct the images being compared. The gain reference images in *Leginon* are also not linearly fitted to intensities obtained with increasing exposure times, as they are in *UCSF Tomo*, but are instead a simple average of a number of images at the same exposure time. *Leginon* also does not rescale quadrant effects from multi-quadrant cameras like *UCSF Tomo*. Within the prediction algorithm itself, separate tracking models are used in *Leginon* in the positive and negative tilt angle ranges to better tolerate differences in goniometer behavior. In order to facilitate recovery from unusual stage movements, a new option for resetting the parameters used in the tracking model at any time during the tilt-series has been included. Finally, while the tilt axis angle and offset are pre-calibrated constants in *UCSF Tomo*'s stage-movement modeling algorithm, *Leginon* offers the option of refining these dynamically after each image. When dynamically fitting the tilt axis angle and offset, all the shifts in the tilt-series are used to (over-)determine the more complicated prediction model. If, however, dynamic fitting is disabled, only the previous four image shifts are used in the prediction so it can be more sensitive to local variations.

5. Data archiving

One of the advantages of automation is that it allows individuals and groups to store, organize, and query comprehensive “meta-data” about the data collection process as well as the actual images. *Leginon*, for instance, stores nearly every setting and parameter available about the microscope and the images in a MySQL database. MySQL communication clients are available for a number of platforms and software packages, and tools for interacting with the *Leginon* database from any web browser with appropriate access have been written in Python, PHP, and Matlab (The MathWorks Inc.).

As part of this effort, several new web tools have been added to help users monitor tilt-series acquisition and analyze patterns. These are accessed from and organized by a “main” tomography page, which displays all the sessions for which data is available, each of the tilt-series in those sessions, and detailed information about whichever tilt-series is currently selected (Fig. 3). The information shown includes a set of thumbnail images of the tilt-series to verify if the tilt-series is progressing correctly; graphs of the actual and predicted stage trajectories in x, y, and z; and the mean counts per image. A link is provided to display the change in energy filter slit position and dose calibration over the entire session. Another link uses a PHP script to dynamically generate an “MRC”-format stack from the individual MRC images stored by *Leginon* and download it to the users' workstation.

The persistence of and ease of access to data on almost all aspects of data collection makes the *Leginon* database a valuable tool not only for development of better methods but also for diagnosing problems with the microscope. It has proven useful, for example, to retrieve and/or graph the x and y displacements measured during the tilt-series to evaluate the performance of the trajectory prediction algorithm and test alternative strategies. The trajectory of the specimen also provides insight into the performance of the goniometer. Patterns and shifts can be detected as they gradually or suddenly develop, guiding service and minimizing the time lost to failed data collection when the instrument performance is poor.



Fig. 3. Screenshot of MSI Tomography's database web interface tool. This tool allows users to review previous tilt-series or check on the progress of current sessions from any web browser with access to the *Legion* webserver. A drop-down list at the top allows the user to select the data collection session of interest (here "08nov18a"). The various tilt-series collected during that session then appear as a list on the left. After a specific tilt-series is selected (here "11/18/08 16:26:39"), a row of thumbnail images display snapshots of the tilt-series at intervals throughout the tilt range to give the observer a visual sense of the target and how well it was tracked. Here, the target was a slender bacterial cell (long grey streak emerging from the bottom left corner of the images) whose tip is suspended over a circular hole in the carbon film. The movement of the specimen and the performance of the tracking algorithm are plotted below the thumbnails. Graphs for z-height and image mean value also appear lower on the web page, but are not shown here for lack of space. Links to download an assembled MRC-format image stack ("Download MRC stack") and display additional graphs of the energy slit and beam intensity change ("Summary") are also provided. In the plots, the green "feature" curves show how far away from the center of each image the target appeared (i.e., the tracking errors, in pixels, right vertical axes). The blue "prediction" curves report where the specimen was expected to be within the column as each image was taken, saved during the tilt-series as the total beam shift applied (in microns, left vertical axes). The orange "position" curves show the *actual* trajectory the specimen traversed (the sum of the beam shifts applied before the image was taken and the actual location of the target observed in the image, in microns, left vertical axes). The x-axes correspond to the tilt angle in degrees. Because the stage is physically re-centered on the target between the first and second halves of the tilt-series and two "untitled" images are recorded, there are actually two "0°" points plotted next to each other in the center. In order to understand the relationship of the curves and the order of calculations and events, details of the first few operations shown will be described with reference to the "y-axis" plot, since the changes are large enough there to be followed in the graph. Before the tilt-series is acquired, the target is approximately centered on the CCD using stage shifts. The remaining fine shift needed to precisely re-center the target is done with beam shifts, and that shift is plotted as the first "prediction" point at 0° tilt. In this case the initial beam shift applied in the y-direction was 0.17 μm (blue curve, rightmost of the two adjacent 0° points). This value was considered the "actual" starting position of the specimen in the column as well, so the orange "position" curve begins at the same point. In the first image, the target was assumed to be correctly centered, so the green "feature" curve begins at exactly zero. Before the second image was taken, no predictions were made about how the specimen would move, so no additional beam shifts were applied, and the prediction curve for the 1° image remains flat at 0.17 μm . After the second image was recorded, the position of the specimen in the image was measured by cross-correlation, and its deviation from the center (33 pixels) was plotted as the first tracking error ("feature" curve at 1°). The actual specimen position curve was therefore plotted an equivalent distance (0.03 μm) higher (note sign conventions are such that beam shift corrections oppose observed shifts in the images), at 0.2 μm . Given this first shift, a prediction was then made about where the specimen would be after the grid was tilted to 2°. The result was applied as a (modified) beam shift and plotted as the "prediction" (0.22 μm). This prediction proved largely correct, as the target then appeared just 2 pixels above the center of the 2° image (green curve). As a result, the "position" was also plotted as 0.22 μm (orange curve, 2°). Based on this trend, a further beam shift of 0.03 μm was applied before the 3° image was taken (blue curve now at 0.25°, 3°), and so forth until the 60° image. The specimen can be seen to have moved quite steadily "up" in the y direction, leading to very small tracking errors, until it unexpectedly hooked "down" between 34° and 39°, leading to tracking errors of first 30 pixels in one direction (34° and 35° images) and then 42 pixels in the other (37° image) before the tracking re-stabilized. After the first half of the tilt-series was finished, the stage was rotated back to 0° and the target was roughly re-centered with stage shifts. The remaining fine shifts needed to precisely re-center the target before the second half of the tilt-series appear as the initial beam shifts (0.22 μm , blue and orange curves, leftmost 0° point). Only very slight tracking errors were seen in the second half of the tilt-series (negative tilt angles).

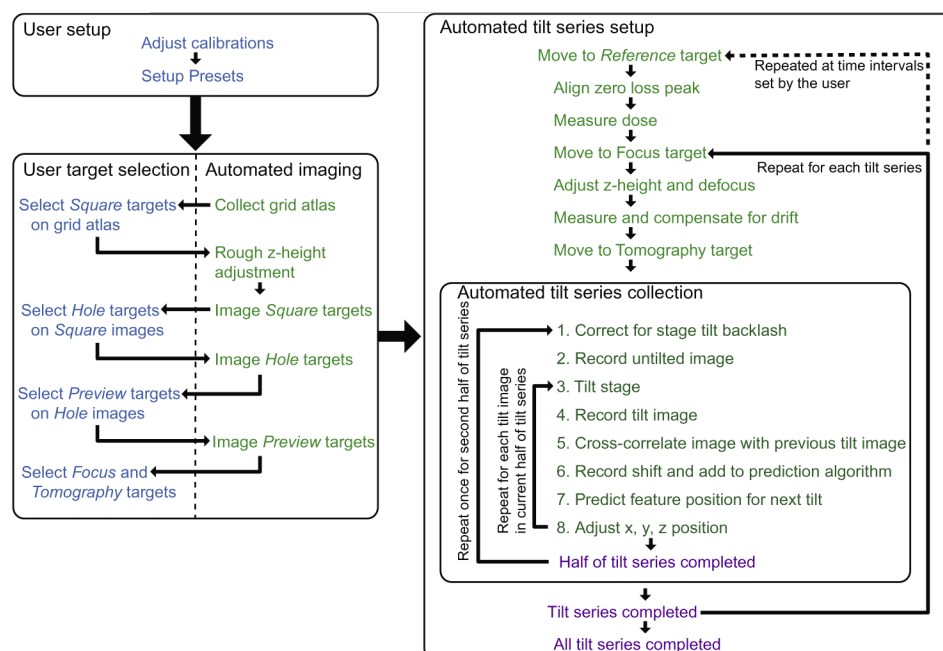


Fig. 4. Diagram of the workflow for MSI Tomography. User tasks are shown in blue, automated tasks are shown in green, and completed tasks are shown in purple. User setup usually takes 30–60 min. Target selection is semi-automated, and takes one to several hours depending on the nature and number of targets desired. An atlas of the grid is automatically acquired, for instance, and then the user selects a small number of good squares for further inspection. This pattern is repeated at higher magnifications until all the targets are selected and corresponding focus positions are marked. The rest of the process is fully automated. Each tilt-series takes 30–60 min depending on the number of tilt angles.

Table 2

Example *Leginon* tomography session results, parameters and performance. For further characterization of *Leginon*'s ability to re-center targets and UCSF Tomo's ability to track them during tilt-series, please also see Zheng et al. (2004), Zheng et al. (2007), Potter et al. (1999), and Suloway et al. (2005).

Session name	Number of tilt-series	Tilt range (°)	Tilt increment (°)	Pixel size (nm)	Defocus (μm)	Image size (pixels)	Binning (pixels)	Time (h)	Mean re-centering (pixels)	Mean tracking error (pixels)
08sep10a	71	–60 to 60	1	0.96	–10	2048 × 2048	2	~32	74 ± 54 (n = 65)	13 ± 26 (n = 7074)
08oct30a	40	–60 to 60	1	0.64	–12	2048 × 2048	2	~22	164 ± 65 (n = 40)	13 ± 24 (n = 4880)
09jan04b	42	–60 to 60	1	0.96	–10	2048 × 2048	2	~22	84 ± 46 (n = 42)	12 ± 17 (n = 3538)

6. Example results

Fully automated, sequential tilt-series acquisition with *Leginon* has now been performed by two labs, namely the Jensen lab at the California Institute of Technology (<http://www.jensenlab.caltech.edu/>) and the National Resource for Automated Molecular Microscopy at the Scripps Research Institute (NRAMM, <http://nramm.scripps.edu>). Most of the tomography data used in Henderson et al. (2007) and in Sharma et al. (2009) were taken with *Leginon*. At Caltech, *Leginon* has been used on both an FEI Tecnai T12 and F30 Polara (FEI Company) since early 2006 to record thousands of tilt-series of viruses and cells in vitreous ice. *Leginon* allowed over 600 tilt-series to be attempted on the Caltech Polara in just 1 month (December 2008). *Leginon* has collected tilt-series with pixel sizes from 2.5–12 Å and at 3–12 μm underfocus. *Leginon* tomography sessions have now run continuously for over 60 h, including over extended holiday weekends, with the only human interaction necessary being to refill the liquid nitrogen Dewar flask on the microscope every ~24 h. Because on the Polara the average time per image of a tilt-series is about 30 s (not counting the over-

head of energy filter adjustment, dose calibration, drift management, and focusing); and tilt-series from ±70° at 2° increments (typical for viruses) and ±60° at 1° increments (typical for cells) take 30 min and 1 h, respectively; such multi-day sessions typically produce 50 or more tilt-series (Table 2). Several examples of data collection at Caltech are presented in Table 2 including mean re-centering and tracking errors for each session. Compared to UCSF Tomo, the tilt-series produced by *Leginon* are of equal quality and take a similar amount of time, but are easier to obtain, since the microscope records one after another through evenings and weekends without user intervention.

Leginon is distributed under the Apache License, Version 2.0 and is available for download from the website <http://leginon.org>.

Acknowledgments

We thank Drs. Bill Tivol, Prabha Dias, Jane Ding, Lu Gan, Ariane Briegel, Elizabeth Wright, Zhou Li and Alasdair McDowall for advice and testing during the development of *MSI Tomography*. This work was supported in part by NIH Grants R01 AI067548 and

P50 GM082545 and DOE Grant DE-FG02-04ER63785 to GJJ, a Searle Scholar Award to GJJ, the Beckman Institute at Caltech, and gifts to Caltech from the Gordon and Betty Moore Foundation and Agouron Institute. *Leginon* development at the National Resource for Automated Molecular Microscopy is supported by the National Institutes of Health through the National Center for Research Resources' P41 program (RR17573 and R01 RR023093).

References

- Briegleb, A., Ding, H.J., Li, Z., Werner, J., Gitai, Z., Dias, D.P., Jensen, R.B., Jensen, G.J., 2008. Location and architecture of the *Caulobacter crescentus* chemoreceptor array. *Mol. Microbiol.* 69, 30–41.
- DeRosier, D.J., Moore, P.B., 1970. Reconstruction of three-dimensional images from electron micrographs of structures with helical symmetry. *J. Mol. Biol.* 52, 355–369.
- Hart, R.G., 1968. Electron microscopy of unstained biological material: the polytropic montage. *Science* 159, 1464–1467.
- Henderson, G.P., Gan, L., Jensen, G.J., 2007. 3D ultrastructure of *O. tauri*: electron cryotomography of an entire eukaryotic cell. *PLoS One* 2, e749.
- Hoppe, W., Langer, R., Knesch, G., Poppe, C., 1968. Protein-Kristallstrukturanalyse mit Elektronenstrahlen. *Naturwissenschaften* 55, 333–336.
- Jensen, G.J., Briegleb, A., 2007. How electron cryotomography is opening a new window onto prokaryotic ultrastructure. *Curr. Opin. Struct. Biol.* 17, 260–267.
- Koster, A.J., van den Bos, A., van der Mast, K.D., 1987. An autofocus method for a TEM. *Ultramicroscopy* 21, 209–222.
- Koster, A.J., Chen, H., Sedat, J.W., Agard, D.A., 1992. Automated microscopy for electron tomography. *Ultramicroscopy* 46, 207–227.
- Lei, J., Frank, J., 2005. Automated acquisition of cryo-electron micrographs for single particle reconstruction on an FEI Tecnai electron microscope. *J. Struct. Biol.* 150, 69–80.
- Li, Z., Trimble, M.J., Brun, Y.V., Jensen, G.J., 2007. The structure of FtsZ filaments in vivo suggests a force-generating role in cell division. *EMBO J.* 26, 4694–4708.
- Lučić, V., Förster, F., Baumeister, W., 2005. Structural studies by electron tomography: from cells to molecules. *Annu. Rev. Biochem.* 74, 833–865.
- Marsh, B.J., 2005. Lessons from tomographic studies of the mammalian Golgi. *Biochim. Biophys. Acta* 1744, 273–292.
- Mastronarde, D.N., 2005. Automated electron microscope tomography using robust prediction of specimen movements. *J. Struct. Biol.* 152, 36–51.
- McIntosh, R., Nicastro, D., Mastronarde, D., 2005. New views of cells in 3D: an introduction to electron tomography. *Trends Cell Biol.* 15, 43–51.
- Murphy, G.E., Jensen, G.J., 2005. Electron cryotomography of the *E. coli* pyruvate and 2-oxoglutarate dehydrogenase complexes. *Structure* 13, 1765–1773.
- Nickell, S., Förster, F., Linares, A., Net, W.D., Beck, F., Hegerl, R., Baumeister, W., Plitzko, J.M., 2005. TOM software toolbox: acquisition and analysis for electron tomography. *J. Struct. Biol.* 149, 227–234.
- Oostergetel, G.T., Keegstra, W., Brissos, A., 1998. Automation of specimen selection and data acquisition for protein electron crystallography. *Ultramicroscopy* 74, 47–59.
- Potter, C.S., Chu, H., Frey, B., Green, C., Kisseberth, N., Madden, T.J., Miller, K.L., Nahrstedt, K., Pulokas, J., Reilein, A., Tchong, D., Weber, D., Carragher, B., 1999. *Leginon*: a system for fully automated acquisition of 1000 electron micrographs a day. *Ultramicroscopy* 77, 153–161.
- Sharma, J., Chhabra, R., Cheng, A., Brownell, J., Liu, Y., Yan, H., 2009. Control of self-assembly of DNA tubules through integration of gold nanoparticles. *Science* 323, 112–116.
- Shi, J., Williams, D.R., Stewart, P.L., 2008. A Script-Assisted Microscopy (SAM) package to improve data acquisition rates on FEI Tecnai electron microscopes equipped with Gatan CCD cameras. *J. Struct. Biol.* 164, 166–169.
- Suloway, C., Pulokas, J., Fellmann, D., Cheng, A., Guerra, F., Quispe, J., Stagg, S., Potter, C.S., Carragher, B., 2005. Automated molecular microscopy: the new *Leginon* system. *J. Struct. Biol.* 151, 41–60.
- Yoshioka, C., Pulokas, J., Fellmann, D., Potter, C.S., Milligan, R.A., Carragher, B., 2007. Automation of random conical tilt and orthogonal tilt data collection using feature-based correlation. *J. Struct. Biol.* 159, 335–346.
- Zhang, P., Beatty, A., Milne, J.L., Subramaniam, S., 2001. Automated data collection with a Tecnai 12 electron microscope: applications for molecular imaging by cryomicroscopy. *J. Struct. Biol.* 135, 251–261.
- Zheng, Q.S., Braumfeld, M.B., Sedat, J.W., Agard, D.A., 2004. An improved strategy for automated electron microscopic tomography. *J. Struct. Biol.* 147, 91–101.
- Zheng, S.Q., Keszthelyi, B., Branlund, E., Lyle, J.M., Braumfeld, M.B., Sedat, J.W., Agard, D.A., 2007. UCSF tomography: an integrated software suite for real-time electron microscopic tomographic data collection, alignment, and reconstruction. *J. Struct. Biol.* 157, 138–147.